

Report generated by: 15:56:50 2003
Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

the number of results printed by chance to have a
value of equal to the number of the result being printed,
if the analysis of the results is not a distribution.

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

Report generated by: 15:56:50 2003

ALIGNMENTS

Report generated by: 15:56:50 2003

Genomic map of the 15.4 Mb region
15.4 Mb region of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region

Genomic map of the 15.4 Mb region



software version 4.10.4508
Apr 01 2003 06:45:00 AM

GenBank accession

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

Accession: 1455142247 (2003)

score greater than or equal to the score of the first hit, printed,
and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Match	Length	DB	ID	Description
1	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
2	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
3	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
4	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
5	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
6	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
7	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
8	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
9	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
10	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
11	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
12	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
13	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
14	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
15	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
16	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
17	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
18	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
19	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
20	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
21	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
22	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
23	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
24	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
25	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
26	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
27	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
28	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
29	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
30	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
31	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
32	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
33	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
34	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
35	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
36	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
37	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
38	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
39	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
40	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
41	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
42	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
43	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
44	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein
45	1273.8	75.6	1636	4	DOGGP80	M6251 Dog glycoprotein

ALIGNMENTS

RESULT 1

DOGGP80

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

DOGGP80
Dog glycoprotein
M5521 M5521
M5521.1 G116194
glycoprotein
Canis familiaris
Canis familiaris
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1. Bases 1 to 1636
Appel, D., Haase, W., Mann, K., Weller, A., and Koch, R.
Molecular cloning of a dog glycoprotein complex secreted by


```

* 7177 7476: gap of unknown length
* 7477 8725: contig of 1249 bp in length
* 8726 8825: gap of unknown length
* 8825 10188: contig of 1363 bp in length
* 10188 10788: gap of unknown length
* 10788 11454: contig of 1166 bp in length
* 11454 11555: gap of unknown length
* 11555 13173: contig of 1519 bp in length
* 13173 13273: gap of unknown length
* 13273 14432: contig of 1159 bp in length
* 14432 14532: gap of unknown length
* 14532 15798: contig of 1266 bp in length
* 15798 15898: gap of unknown length
* 15898 16373: contig of 1031 bp in length
* 16373 16779: gap of unknown length
* 16779 18647: contig of 1569 bp in length
* 18647 18747: gap of unknown length
* 18747 18848: contig of 1037 bp in length
* 18848 19344: gap of unknown length
* 19344 21319: contig of 1374 bp in length
* 21319 21418: gap of unknown length
* 21418 21577: contig of 1492 bp in length
* 21577 23177: gap of unknown length
* 23177 24188: contig of 1187 bp in length
* 24188 24288: gap of unknown length
* 24288 25690: contig of 1112 bp in length
* 25690 26690: gap of unknown length
* 26690 27143: contig of 1463 bp in length
* 27143 27543: gap of unknown length
* 27543 28343: contig of 1106 bp in length
* 28343 28443: gap of unknown length
* 28443 29662: contig of 1213 bp in length
* 29662 30520: gap of unknown length
* 30520 31161: contig of 1399 bp in length
* 31161 32770: contig of 1109 bp in length
* 32770 33470: gap of unknown length
* 33470 34107: contig of 1637 bp in length
* 34107 34207: gap of unknown length
* 34207 35394: contig of 1187 bp in length
* 35394 36494: gap of unknown length
* 36494 36896: contig of 1402 bp in length
* 36896 38478: gap of unknown length
* 38478 38599: contig of 1363 bp in length
* 38599 38459: gap of unknown length
* 38459 39572: contig of 1233 bp in length
* 39572 39742: gap of unknown length
* 39742 40000: contig of 1261 bp in length
* 40000 41253: gap of unknown length
* 41253 43161: contig of 2068 bp in length
* 43161 43261: gap of unknown length
* 43261 44459: contig of 1199 bp in length
* 44459 44577: gap of unknown length
* 44577 45277: contig of 1268 bp in length
* 45277 45727: gap of unknown length
* 45727 47186: contig of 1259 bp in length
* 47186 47286: gap of unknown length
* 47286 49071: contig of 1695 bp in length
* 49071 50371: gap of unknown length
* 50371 50471: contig of 1270 bp in length
* 50471 51421: gap of unknown length
* 51421 51855: contig of 1414 bp in length
* 51855 51955: gap of unknown length
* 51955 53089: contig of 1134 bp in length
* 53089 53189: gap of unknown length
* 53189 54954: contig of 1765 bp in length
* 54954 55494: gap of unknown length
* 55494 56491: contig of 1437 bp in length
* 56491 56591: gap of unknown length
* 56591 58292: contig of 1701 bp in length
* 58292 58392: gap of unknown length
* 58392 59309: contig of 1517 bp in length
* 59309 60009: gap of unknown length

```

```

* 60010 61430: contig of 1428 bp in length
* 61430 61577: gap of unknown length
* 61577 64323: contig of 2496 bp in length
* 64323 64423: gap of unknown length
* 64423 65754: contig of 1572 bp in length
* 65754 65804: gap of unknown length
* 65804 67677: contig of 1873 bp in length
* 67677 67777: gap of unknown length
* 67777 69446: contig of 1669 bp in length
* 69446 69546: gap of unknown length
* 69546 70900: contig of 1354 bp in length
* 70900 71000: gap of unknown length
* 71000 71231: contig of 231 bp in length
* 71231 73331: gap of unknown length
* 73331 74331: contig of 1031 bp in length
* 74331 74431: gap of unknown length
* 74431 75343: contig of 2072 bp in length
* 75343 75731: gap of unknown length
* 75731 78509: contig of 1875 bp in length
* 78509 78609: gap of unknown length
* 78609 80611: contig of 2002 bp in length
* 80611 82711: gap of unknown length
* 82711 84492: contig of 1698 bp in length
* 84492 85443: contig of 1041 bp in length
* 85443 85543: contig of 1001 bp in length
* 85543 85643: contig of 1001 bp in length
* 85643 85743: contig of 1001 bp in length
* 85743 85843: contig of 1001 bp in length
* 85843 85943: contig of 1001 bp in length
* 85943 86043: contig of 1001 bp in length
* 86043 86143: contig of 1001 bp in length
* 86143 86243: contig of 1001 bp in length
* 86243 86343: contig of 1001 bp in length
* 86343 86443: contig of 1001 bp in length
* 86443 86543: contig of 1001 bp in length
* 86543 86643: contig of 1001 bp in length
* 86643 86743: contig of 1001 bp in length
* 86743 86843: contig of 1001 bp in length
* 86843 86943: contig of 1001 bp in length
* 86943 87043: contig of 1001 bp in length
* 87043 87143: contig of 1001 bp in length
* 87143 87243: contig of 1001 bp in length
* 87243 87343: contig of 1001 bp in length
* 87343 87443: contig of 1001 bp in length
* 87443 87543: contig of 1001 bp in length
* 87543 87643: contig of 1001 bp in length
* 87643 87743: contig of 1001 bp in length
* 87743 87843: contig of 1001 bp in length
* 87843 87943: contig of 1001 bp in length
* 87943 88043: contig of 1001 bp in length
* 88043 88143: contig of 1001 bp in length
* 88143 88243: contig of 1001 bp in length
* 88243 88343: contig of 1001 bp in length
* 88343 88443: contig of 1001 bp in length
* 88443 88543: contig of 1001 bp in length
* 88543 88643: contig of 1001 bp in length
* 88643 88743: contig of 1001 bp in length
* 88743 88843: contig of 1001 bp in length
* 88843 88943: contig of 1001 bp in length
* 88943 89043: contig of 1001 bp in length
* 89043 89143: contig of 1001 bp in length
* 89143 89243: contig of 1001 bp in length
* 89243 89343: contig of 1001 bp in length
* 89343 89443: contig of 1001 bp in length
* 89443 89543: contig of 1001 bp in length
* 89543 89643: contig of 1001 bp in length
* 89643 89743: contig of 1001 bp in length
* 89743 89843: contig of 1001 bp in length
* 89843 89943: contig of 1001 bp in length
* 89943 90043: contig of 1001 bp in length
* 90043 90143: contig of 1001 bp in length
* 90143 90243: contig of 1001 bp in length
* 90243 90343: contig of 1001 bp in length
* 90343 90443: contig of 1001 bp in length
* 90443 90543: contig of 1001 bp in length
* 90543 90643: contig of 1001 bp in length
* 90643 90743: contig of 1001 bp in length
* 90743 90843: contig of 1001 bp in length
* 90843 90943: contig of 1001 bp in length
* 90943 91043: contig of 1001 bp in length
* 91043 91143: contig of 1001 bp in length
* 91143 91243: contig of 1001 bp in length
* 91243 91343: contig of 1001 bp in length
* 91343 91443: contig of 1001 bp in length
* 91443 91543: contig of 1001 bp in length
* 91543 91643: contig of 1001 bp in length
* 91643 91743: contig of 1001 bp in length
* 91743 91843: contig of 1001 bp in length
* 91843 91943: contig of 1001 bp in length
* 91943 92043: contig of 1001 bp in length
* 92043 92143: contig of 1001 bp in length
* 92143 92243: contig of 1001 bp in length
* 92243 92343: contig of 1001 bp in length
* 92343 92443: contig of 1001 bp in length
* 92443 92543: contig of 1001 bp in length
* 92543 92643: contig of 1001 bp in length
* 92643 92743: contig of 1001 bp in length
* 92743 92843: contig of 1001 bp in length
* 92843 92943: contig of 1001 bp in length
* 92943 93043: contig of 1001 bp in length
* 93043 93143: contig of 1001 bp in length
* 93143 93243: contig of 1001 bp in length
* 93243 93343: contig of 1001 bp in length
* 93343 93443: contig of 1001 bp in length
* 93443 93543: contig of 1001 bp in length
* 93543 93643: contig of 1001 bp in length
* 93643 93743: contig of 1001 bp in length
* 93743 93843: contig of 1001 bp in length
* 93843 93943: contig of 1001 bp in length
* 93943 94043: contig of 1001 bp in length
* 94043 94143: contig of 1001 bp in length
* 94143 94243: contig of 1001 bp in length
* 94243 94343: contig of 1001 bp in length
* 94343 94443: contig of 1001 bp in length
* 94443 94543: contig of 1001 bp in length
* 94543 94643: contig of 1001 bp in length
* 94643 94743: contig of 1001 bp in length
* 94743 94843: contig of 1001 bp in length
* 94843 94943: contig of 1001 bp in length
* 94943 95043: contig of 1001 bp in length
* 95043 95143: contig of 1001 bp in length
* 95143 95243: contig of 1001 bp in length
* 95243 95343: contig of 1001 bp in length
* 95343 95443: contig of 1001 bp in length
* 95443 95543: contig of 1001 bp in length
* 95543 95643: contig of 1001 bp in length
* 95643 95743: contig of 1001 bp in length
* 95743 95843: contig of 1001 bp in length
* 95843 95943: contig of 1001 bp in length
* 95943 96043: contig of 1001 bp in length
* 96043 96143: contig of 1001 bp in length
* 96143 96243: contig of 1001 bp in length
* 96243 96343: contig of 1001 bp in length
* 96343 96443: contig of 1001 bp in length
* 96443 96543: contig of 1001 bp in length
* 96543 96643: contig of 1001 bp in length
* 96643 96743: contig of 1001 bp in length
* 96743 96843: contig of 1001 bp in length
* 96843 96943: contig of 1001 bp in length
* 96943 97043: contig of 1001 bp in length
* 97043 97143: contig of 1001 bp in length
* 97143 97243: contig of 1001 bp in length
* 97243 97343: contig of 1001 bp in length
* 97343 97443: contig of 1001 bp in length
* 97443 97543: contig of 1001 bp in length
* 97543 97643: contig of 1001 bp in length
* 97643 97743: contig of 1001 bp in length
* 97743 97843: contig of 1001 bp in length
* 97843 97943: contig of 1001 bp in length
* 97943 98043: contig of 1001 bp in length
* 98043 98143: contig of 1001 bp in length
* 98143 98243: contig of 1001 bp in length
* 98243 98343: contig of 1001 bp in length
* 98343 98443: contig of 1001 bp in length
* 98443 98543: contig of 1001 bp in length
* 98543 98643: contig of 1001 bp in length
* 98643 98743: contig of 1001 bp in length
* 98743 98843: contig of 1001 bp in length
* 98843 98943: contig of 1001 bp in length
* 98943 99043: contig of 1001 bp in length
* 99043 99143: contig of 1001 bp in length
* 99143 99243: contig of 1001 bp in length
* 99243 99343: contig of 1001 bp in length
* 99343 99443: contig of 1001 bp in length
* 99443 99543: contig of 1001 bp in length
* 99543 99643: contig of 1001 bp in length
* 99643 99743: contig of 1001 bp in length
* 99743 99843: contig of 1001 bp in length
* 99843 99943: contig of 1001 bp in length
* 99943 100043: contig of 1001 bp in length

```




17 108 5.7 716 5 Q9RUC6
18 107 5.7 798 5 Q01743
19 107 5.6 432 10 Q8182
20 107 5.6 7231 5 Q9MT8
21 106 5.5 696 4 Q8755
22 106 5.5 1738 5 Q8323
23 104 5.4 1833 2 Q9L89
24 104 5.4 1623 5 Q9307
25 103.5 5.4 363 11 Q9EQ9
26 103.5 5.4 625 6 Q9217
27 103.5 5.4 2867 5 Q9X23
28 103.5 5.4 3160 4 Q9X28
29 103.5 5.4 531 4 Q9X29
30 103 5.4 543 10 Q91X2
31 103 5.4 774 16 Q9215
32 103 5.4 1780 4 Q9QF5
33 103 5.4 1820 4 Q9Q89
34 102 5.3 1266 4 Q9NT1
35 101 5.3 990 5 Q9SV5
36 100.5 5.3 439 10 Q9FJ7
37 100.5 5.3 1173 4 Q9U86
38 100.5 5.3 1312 4 Q92878
39 100.5 5.3 1318 4 Q93254
40 100.5 5.3 1728 10 Q9LH2
41 100 5.2 651 12 Q9DVI2
42 100 5.2 1217 4 Q60464
43 100 5.2 1217 11 Q9CUS1
44 99.5 5.2 548 2 Q9A74
45 99.5 5.2 1388 4 Q9NS87

Query Match 76.9% Score 1490.5; DP 6; Length 440

ALIGNMENTS

RESULT 1
Q29482 PRELIMINARY; PRT: 448 AA.
AC Q29482; T1EMHRL61: 01, Created
DT 01-NOV-1996 (T1EMHRL61: 01, Last sequence update)
DT 01-NOV-1996 (T1EMHRL61: 01, Last sequence update)
DT 01-DEC-2001 (T1EMHRL61: 13, Last annotation update)
Clusterin precursor.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
OF Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI TaxID 9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STEAM-OLIVER; TISSUE-TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Craik R.D., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
Clusterin."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO ELISA, MEMBRANES,
ANT HYPERTENSIVE FACTORS. IT HAS BEEN ASSOCIATED WITH INFLAMED
CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTI-PARALLEL DISULFIDE-LINKED HEIF OLIMER (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR ENBL: L46797; AAA80311.1; 70
DR InterPro: I2500753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; CLA; 1.
DR SMART: SM00030; CLA; 1.
DR PROSITE: PS00492; CLUSTERIN 1; 1.
DR PROSITE: PS00493; CLUSTERIN 2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 448 CLUSTERIN.
CC SEQUENCE 448 AA, 5116 MW, 137.60 kDa, 10.7 kDa

Query Match 76.9% Score 1490.5; DP 6; Length 440

```

Best Local Similarity 84.18, Prod. No. 1140-102;
Matches 296, Conservation 21, Mismatch 10, Indels 4, Gaps 0

QY 1 MPTLLVGLLWNSGVLNCHWCTELTETVTVVDFIPACVDTPTTTE 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MPTLLVGLLWNSGVLNCHWCTELTETVTVVDFIPACVDTPTTTE 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 41 TNSHFFSLTHFAPPPFAVHTYLSKAPFAVHTYLSKAPFAVHTYLSK 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 CSHFFPFLSTHTAPPPFAVHTYLSKAPFAVHTYLSKAPFAVHTYLSK 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 101 MPTVATVPCSTGLVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 MPTVATVPCSTGLVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 191 RASSIMLELQVAFIRFQAALHHSFSSQVRR EFTFNKHFTASIMHFTTQWTF 239
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 191 RASSIMLELQVAFIRFQAALHHSFSSQVRR EFTFNKHFTASIMHFTTQWTF 239
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 249 HAMPSHCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 249 HAMPSHCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 299 CQERBELVLYCQCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 299 CQERBELVLYCQCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 301 CQERBELVLYCQCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 CQERBELVLYCQCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
QYAK94
ID QYAK94 PRELIMINARY, FRT, 372 AA.
AC QYAK94;
DT 01-OCT-2000 (TREMBlrel, 15, Created)
DT 01-OCT-2000 (TREMBlrel, 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel, 19, Last annotation update)
DE Clusterin isoform 2.
OS Mus musculus (Mouse).
OC Eukaryota, Euteleostomi, Chordata, Cladocera, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
PC STRAIN BALB/c, ISSUS-UTERUS,
PA You K.H., Jeon J.H.;
RT Identification of Truncated Cgf 2 lacking a Signal Peptide for
RT Secretion".
PL Submitted (MAR 2000) to the EMBL/GenBank/TrEMBL database.
CC 1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT BEING ABLE TO BIND TO SERUM MEMBRANES,
CC AND HIEROGLYPHS. IN ADDITION, IT HAS BEEN ASSOCIATED WITH FETTERED
CC CELL DEATH (BY SIMILARITY).
CC 1- SUBUNIT: ANTIPARALLEL DISULFIDE LINKED HETERODIMER (BY
CC SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC EMBL: AF248658; AAF67195.1;
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00030; C1b; 1.
CC SMART: SM00030; C1b; 1.
CC PROSITE: PS00492; CLUSTERIN 1; 1.
CC PROSITE: PS00493; CLUSTERIN 2; 1.
CC GlycoProtein.
CC SEQUENCE 372 AA; A1047E48520FE8 CPC64;

Query Match 49.62, Score 972, E-11, Length 322;
Best Local Similarity 62.14, Prod. No. 100-04;
Matches 161, Conservation 26, Mismatch 59, Indels 4, Gaps 0

QY 79 EYEDALNKELEMLELALVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 EYEDALNKELEMLELALVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 129 CVELFNLCSTFFTNKNGELTLENTNCTHVALDWTLEIFASINELTFQETTF 197
| | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 3
QYAK94
ID QYAK94 PRELIMINARY, FRT, 295 AA.
AC QYAK94;
DT 01-JAN-1998 (TREMBlrel, 05, Created)
DT 01-JAN-1998 (TREMBlrel, 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel, 19, Last annotation update)
DE Clusterin.
OS Rattus norvegicus (Rat).
OC Eukaryota, Euteleostomi, Chordata, Cladocera, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
PC STRAIN WISTAR, TIGROE CEMINAL VESICLE;
PC MEDLINE: 98163873; PubMed: 9503143;
PA Izawa M.;
RT Identification of a transcript prediction as alternative form of
RT silfoid glycoprotein? Clusterin in rat testes."
RL Biochem. Mol. Biol. Int. 44:9-19(1998)
CC 1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT BEING ABLE TO BIND TO SERUM MEMBRANES,
CC AND HIEROGLYPHS. IN ADDITION, IT HAS BEEN ASSOCIATED WITH FETTERED
CC CELL DEATH (BY SIMILARITY).
CC 1- SUBUNIT: ANTIPARALLEL DISULFIDE LINKED HETERODIMER (BY
CC SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC EMBL: D11339; BAA21775.1;
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SMART; C1a; 1.
CC SMART: SMART; C1b; 1.
CC PROSITE: PS00492; CLUSTERIN 1; 1.
CC PROSITE: PS00493; CLUSTERIN 2; 1.
CC GlycoProtein.
CC SEQUENCE 295 AA; A1047E48520FE8 CPC64;

Query Match 42.03, Score 936, E-11, Length 285;
Best Local Similarity 51.28, Prod. No. 110-62;
Matches 177, Conservation 21, Mismatch 47, Indels 2, Gaps 0

QY 52 MALLALQVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 52 MALLALQVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 100 WNNPFTICLITPSQCTHATVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 100 WNNPFTICLITPSQCTHATVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 121 HRPFLHTLRLRLMLHFLHFLHFLHFLHFLHFLHFLHFLHFLHFLHFLHFLH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 HRPFLHTLRLRLMLHFLHFLHFLHFLHFLHFLHFLHFLHFLHFLHFLHFLH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 200 EFTFNKHFTASIMHFTTQWTF EFTFNKHFTASIMHFTTQWTF EFTFNKHFTASIMHFTTQWTF 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 200 EFTFNKHFTASIMHFTTQWTF EFTFNKHFTASIMHFTTQWTF EFTFNKHFTASIMHFTTQWTF 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 181 DFLKSGNDPTVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 DFLKSGNDPTVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCHVCH 17
| | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
RESULT 5
P79329
ID P79329 PRELIMINARY; PST; 176 AA.
AC P79329;
DT 01-MAY-1997 (TrEMBLrel. 01, Created.
DT 01-MAY-1997 (TrEMBLrel. 01, Last sequence update.
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update.
DE Clusterin (Fragment).
GN SGP-2.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Primata; Platyrrhini; Platyrrhini; Callitrichidae; Callitrichinae.
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrichinae.
OX NCBI_TaxID=9483;
RN [1]
SEQUENCE FROM N.A.
FP TISSUE-TESTIS;
RC Gaughan J., Saunders P.T.K.;
RL Submitted (NCBI) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBMIT: AMERICAN;
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL: Y09532 CAA70724.1;
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; Cld; 1.
DR SMART: SM00030; Cld; 1.
DR SMART: PS00493; CLUSTERIN_2; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
RW Glycoprotein.
FT NON_TER 176 176
SQ SEQUENCE 176 AA; 4086 MW; 416A1A9421B5BF9 F064;

Query Match 31.0%; Score 592; DP 6; Length 176;
Best local similarity 69.1%; Pred. No. 44-37;
Matches 123; Conservative 15; Mismatches 14; Indels 4; Gaps 2;

QY 153 TDEETDCELENTFTHALLVWLLIPLAAGMDEFCALIFPAAPRPPRPPRPPRPP 212
DB 1 MNGDRDILLNDFQOTHCDCGCGGFFRASSMELGCGFFRPPRPPRPPRPPRPP 60

QY 213 KRRTFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNK 272
DB 61 RFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFR 119

QY 273 F RNCQDVAQVETVNTSTGLKMKQCKRCKRCKRCKRCKRCKRCKRCKRCKR 176
DB 120 ESTGRGELFAVEPEFRNSPDELMAVCCYRPELVDSCTPR 166

RESULT 6
Q9ERP1
ID Q9ERP1 PRELIMINARY; PST; 118 AA.
AC Q9ERP1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Clusterin (Fragment).
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Primata; Platyrrhini; Platyrrhini; Callitrichidae; Callitrichinae.
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrichinae.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
FP TISSUE=THYMUS;
RC Park J.H., Park J.S., Ju S.K., Na S.Y., Yoo K.H.;
RL Determination of clusterin mRNA expression of apoptosis-induced rat
RL thymocytes in vivo and in vitro.
RL Submitted (OCT 2000) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
```

```
RESULT 5
P79329
ID P79329 PRELIMINARY; PST; 176 AA.
AC P79329;
DT 01-MAY-1997 (TrEMBLrel. 01, Created.
DT 01-MAY-1997 (TrEMBLrel. 01, Last sequence update.
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update.
DE Clusterin (Fragment).
GN SGP-2.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Primata; Platyrrhini; Platyrrhini; Callitrichidae; Callitrichinae.
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrichinae.
OX NCBI_TaxID=9483;
RN [1]
SEQUENCE FROM N.A.
FP TISSUE-TESTIS;
RC Gaughan J., Saunders P.T.K.;
RL Submitted (NCBI) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBMIT: AMERICAN;
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL: Y09532 CAA70724.1;
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; Cld; 1.
DR SMART: SM00030; Cld; 1.
DR SMART: PS00493; CLUSTERIN_2; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
RW Glycoprotein.
FT NON_TER 176 176
SQ SEQUENCE 176 AA; 4086 MW; 416A1A9421B5BF9 F064;

Query Match 31.0%; Score 592; DP 6; Length 176;
Best local similarity 69.1%; Pred. No. 44-37;
Matches 123; Conservative 15; Mismatches 14; Indels 4; Gaps 2;

QY 153 TDEETDCELENTFTHALLVWLLIPLAAGMDEFCALIFPAAPRPPRPPRPPRPP 212
DB 1 MNGDRDILLNDFQOTHCDCGCGGFFRASSMELGCGFFRPPRPPRPPRPPRPP 60

QY 213 KRRTFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNK 272
DB 61 RFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFRFR 119

QY 273 F RNCQDVAQVETVNTSTGLKMKQCKRCKRCKRCKRCKRCKRCKRCKRCKR 176
DB 120 ESTGRGELFAVEPEFRNSPDELMAVCCYRPELVDSCTPR 166

RESULT 6
Q9ERP1
ID Q9ERP1 PRELIMINARY; PST; 118 AA.
AC Q9ERP1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Clusterin (Fragment).
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Primata; Platyrrhini; Platyrrhini; Callitrichidae; Callitrichinae.
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrichinae.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
FP TISSUE=THYMUS;
RC Park J.H., Park J.S., Ju S.K., Na S.Y., Yoo K.H.;
RL Determination of clusterin mRNA expression of apoptosis-induced rat
RL thymocytes in vivo and in vitro.
RL Submitted (OCT 2000) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
```


Search completed: April 2, 2003, 13:55:12
Job time : 40 secs

1993 497 1.000-000

... 3:54.00, ... 14 Seconds
... Arguments.

100

[illegible]

index

1000

[illegible]

...

[illegible]

series of results probably chance to have a
contribution to the scientific result being printed,
patients of the ... distribution

NAME IN	IT	DESCRIPTION
445	1	GLUS VAMBA
446	1	GLUS IT
447	1	GLUS BURN
448	1	GLUS HUMAN
449	1	GLUS RAT
450	1	GLUS RAT
451	1	GLUS W. RAT
452	1	GLUS CATIA
453	1	GLUS MEXIC
454	1	GLUS HUMA
455	1	GLUS RAT
456	1	GLUS RAT
457	1	GLUS RAT
458	1	GLUS RAT
459	1	GLUS RAT
460	1	GLUS RAT
461	1	GLUS RAT
462	1	GLUS RAT
463	1	GLUS RAT
464	1	GLUS RAT
465	1	GLUS RAT
466	1	GLUS RAT
467	1	GLUS RAT
468	1	GLUS RAT
469	1	GLUS RAT
470	1	GLUS RAT
471	1	GLUS RAT
472	1	GLUS RAT
473	1	GLUS RAT
474	1	GLUS RAT
475	1	GLUS RAT
476	1	GLUS RAT
477	1	GLUS RAT
478	1	GLUS RAT
479	1	GLUS RAT
480	1	GLUS RAT
481	1	GLUS RAT
482	1	GLUS RAT
483	1	GLUS RAT
484	1	GLUS RAT
485	1	GLUS RAT
486	1	GLUS RAT
487	1	GLUS RAT
488	1	GLUS RAT
489	1	GLUS RAT
490	1	GLUS RAT
491	1	GLUS RAT
492	1	GLUS RAT
493	1	GLUS RAT
494	1	GLUS RAT
495	1	GLUS RAT
496	1	GLUS RAT
497	1	GLUS RAT
498	1	GLUS RAT
499	1	GLUS RAT
500	1	GLUS RAT

```

RESULT 1
CICUS CANFA STANDARD: FET: 448 AA
CICUS CANFA P25473;
AC DT 01-MAY-1992 (Rel: 22, Created)
AC DT 01-MAY-1992 (Rel: 22, Last sequence update)
DT DT 15-JUN-2002 (Rel: 41, Last annotation update)
DE DE Clusterin precursor (glycoprotein A6) [MIM:
GN GN CLU]
Canis familiaris (dog)
Eukaryota; Metazoa; Chordata; Euteleostomi; Verte
Mammalia; Eutheria; Carnivora; Euarchontali; Pa
NCBI_TaxID=9615;
(1)_TAXID=9615;
SEQUENCE FROM N.A.
MEDLINE=9136767; PubMed=2033678;
Hartmann K., Rauch J., Urban J., Parezyk K.,
Appel D., Haase W., Mann K., Welter A., Kocher
"MC-Cellar" clusterin of hsp 90. A glycosylated
kidney cells in vitro and in vivo. A link to
and to the complement cascade."
J. Biol. Chem. 268:19924-19931, 1993.
- FUNCTION: Not yet clear. It is known to b
of tissues and it seems to be able to bin
hydrophobic proteins. It has been associa
death.
- SUBMIT: AMITRAKARER DISULFIDE LINKED HE
- SIMILARITY RELATES TO THE LUSHERIN FAM
This SwissProt entry is copyright © 1993 by
between the Swiss Institute of Bioinformatics
the European Bioinformatics Institute. Where
any by any third party without prior writt
modified and this statement is not removed.
entities requires a license agreement. See th
it send an email to license@lsb.scri.ch).
EMBL: M5525; AAA39846.1;
PIR: A40018; A40018.
Triller-Pozzetti HBB00753; clusterin.
Pfam: PF01693; Clusterin; 1.
SMART: SMO0005; Clu; 1.
SWAT: SMO0005; Clb; 1.
Prosite: PS00492; CLUSTERIN; 1.
PROSITE: PS00493; CLUSTERIN; 1.
GlycoSite: 1.
Signal 1 22
CLUSTERIN.
FEFT CHAIN 23 446
FEFT CHAN 03 006
FEFT CHAN 123 445
FEFT CHAN 123 337
FEFT CHAN 123 301
FEFT DISULFID 113 301
FEFT DISULFID 124 294
FEFT DISULFID 121 291
FEFT DISULFID 123 291
FEFT CARBOHYD 80 80

```

1997


```

CC 1 SIMILARITY. BLASTED TO THE NCBI / MBL FAMILY.
CC 2 THIS SWISS PRO entry is copyright. It is provided through a global effort
CC 3 between the Swiss Institute of Bioinformatics and the EMBL outstation
CC 4 at the European Bioinformatics Institute. There are no restrictions on its
CC 5 use by any public institutions as long as its status is not changed
CC 6 modified and this statement is not removed. Usage by third parties
CC 7 requires a license agreement (see http://www.ebi.ac.uk/infocentre
CC 8 or send an email to license@ebi.ac.uk).
CC 9
CC 10 EMBL; AB000000; AAC00000; 17
CC 11 InterPro: IPR000000; Lipoprotein
CC 12 InterPro: IPR000000; Lipoprotein
CC 13 Pfam: PF00000; Lipoprotein
CC 14 Pfam: PF00000; Lipoprotein
CC 15 KEGG: K00000; Lipoprotein
CC 16 KEGG: K00000; Lipoprotein
CC 17 KEGG: K00000; Lipoprotein
CC 18 KEGG: K00000; Lipoprotein
CC 19 KEGG: K00000; Lipoprotein
CC 20 KEGG: K00000; Lipoprotein
CC 21 KEGG: K00000; Lipoprotein
CC 22 KEGG: K00000; Lipoprotein
CC 23 KEGG: K00000; Lipoprotein
CC 24 KEGG: K00000; Lipoprotein
CC 25 KEGG: K00000; Lipoprotein
CC 26 KEGG: K00000; Lipoprotein
CC 27 KEGG: K00000; Lipoprotein
CC 28 KEGG: K00000; Lipoprotein
CC 29 KEGG: K00000; Lipoprotein
CC 30 KEGG: K00000; Lipoprotein
CC 31 KEGG: K00000; Lipoprotein
CC 32 KEGG: K00000; Lipoprotein
CC 33 KEGG: K00000; Lipoprotein
CC 34 KEGG: K00000; Lipoprotein
CC 35 KEGG: K00000; Lipoprotein
CC 36 KEGG: K00000; Lipoprotein
CC 37 KEGG: K00000; Lipoprotein
CC 38 KEGG: K00000; Lipoprotein
CC 39 KEGG: K00000; Lipoprotein
CC 40 KEGG: K00000; Lipoprotein
CC 41 KEGG: K00000; Lipoprotein
CC 42 KEGG: K00000; Lipoprotein
CC 43 KEGG: K00000; Lipoprotein
CC 44 KEGG: K00000; Lipoprotein
CC 45 KEGG: K00000; Lipoprotein
CC 46 KEGG: K00000; Lipoprotein
CC 47 KEGG: K00000; Lipoprotein
CC 48 KEGG: K00000; Lipoprotein
CC 49 KEGG: K00000; Lipoprotein
CC 50 KEGG: K00000; Lipoprotein
CC 51 KEGG: K00000; Lipoprotein
CC 52 KEGG: K00000; Lipoprotein
CC 53 KEGG: K00000; Lipoprotein
CC 54 KEGG: K00000; Lipoprotein
CC 55 KEGG: K00000; Lipoprotein
CC 56 KEGG: K00000; Lipoprotein
CC 57 KEGG: K00000; Lipoprotein
CC 58 KEGG: K00000; Lipoprotein
CC 59 KEGG: K00000; Lipoprotein
CC 60 KEGG: K00000; Lipoprotein
CC 61 KEGG: K00000; Lipoprotein
CC 62 KEGG: K00000; Lipoprotein
CC 63 KEGG: K00000; Lipoprotein
CC 64 KEGG: K00000; Lipoprotein
CC 65 KEGG: K00000; Lipoprotein
CC 66 KEGG: K00000; Lipoprotein
CC 67 KEGG: K00000; Lipoprotein
CC 68 KEGG: K00000; Lipoprotein
CC 69 KEGG: K00000; Lipoprotein
CC 70 KEGG: K00000; Lipoprotein
CC 71 KEGG: K00000; Lipoprotein
CC 72 KEGG: K00000; Lipoprotein
CC 73 KEGG: K00000; Lipoprotein
CC 74 KEGG: K00000; Lipoprotein
CC 75 KEGG: K00000; Lipoprotein
CC 76 KEGG: K00000; Lipoprotein
CC 77 KEGG: K00000; Lipoprotein
CC 78 KEGG: K00000; Lipoprotein
CC 79 KEGG: K00000; Lipoprotein
CC 80 KEGG: K00000; Lipoprotein
CC 81 KEGG: K00000; Lipoprotein
CC 82 KEGG: K00000; Lipoprotein
CC 83 KEGG: K00000; Lipoprotein
CC 84 KEGG: K00000; Lipoprotein
CC 85 KEGG: K00000; Lipoprotein
CC 86 KEGG: K00000; Lipoprotein
CC 87 KEGG: K00000; Lipoprotein
CC 88 KEGG: K00000; Lipoprotein
CC 89 KEGG: K00000; Lipoprotein
CC 90 KEGG: K00000; Lipoprotein
CC 91 KEGG: K00000; Lipoprotein
CC 92 KEGG: K00000; Lipoprotein
CC 93 KEGG: K00000; Lipoprotein
CC 94 KEGG: K00000; Lipoprotein
CC 95 KEGG: K00000; Lipoprotein
CC 96 KEGG: K00000; Lipoprotein
CC 97 KEGG: K00000; Lipoprotein
CC 98 KEGG: K00000; Lipoprotein
CC 99 KEGG: K00000; Lipoprotein
CC 100 KEGG: K00000; Lipoprotein

```


cc between the Swiss Institute of Bioinformatics and the EMBL outstation
 cc the European Bioinformatics Institute. There are no restrictions on its
 cc use by non-profit institutions as long as its content is in no way
 cc modified and this statement is not removed. Usage by and for commercial
 cc entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 cc or send an email to license@isb-sib.ch).

DR EMBL; AL133357; CAB62414.1;
 DR EMBL; AF010473; AAB65416.1; ALT_INIT.
 DR EMBL; AB028012; BAAB/316.1;
 KW Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1105 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1423 COILED COIL (POTENTIAL).
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1727 AA; 197856 MW; P820RPM69C132644 CRC64;

Query Match 5.4%; Score 102.5; DB 1; Length 1727;
 Best Local Similarity 19.5%; Pred. No. 14;
 Matches 83, Conservation 69, Mismatches 172, Indels 10, Gaps 18;

OY 10 NQVDELNMLTLEL EMTEGVVINEI PVALKG 51
 DB 1203 NPLPDELLELLLEKAVVLENSREYVHHLSHLLNSTSTGSAFEPHIV 1328
 OY 50 VVCTPTTETTHFFPGLTTHFFAPPPPT ALNTPYGVPLPAQVYVTHMY 166
 DB 1329 E ENLIELLELELELELELELELELELELELELELELELELELELE 1679
 OY 107 ALW RECHPLKLTOMKPVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 144
 DB 1380 AAPTYYVTVVPSSEVNTTETETETETETETETETETETETETETET 1433
 OY 145 QNSFFVFWIN GRINLELLEMPQVCHALLWJSEFASIMELFQVFFTFEA 199
 DB 1431 ATTLEHGVVLEHNSHLELLLTFLCEHNGHLLHLEHLEHLEHLEHLEH 1493
 OY 210 LIDHHTDSDSTETFF HHHHFAHHHTTTHHLEN FHHHTFFHHHQA 263
 DB 1494 EIDNE FHWPPPVAVVPTVETETETETETETETETETETETETET 1549
 OY 254 LQAMVNHLLHSEMEFEDHLLQAVVPLSHN SLDLE MEVLEHPE 303
 DB 1550 EEARSAVH AHTHELEHLEFHEMLSVTFVPHESISSTSTAGTTPVCHWE 1608
 OY 304 EILSVNCEHNAQVLEF ELNLSGLA EHTLEHFAVAVLPEDYVHV 363
 DB 1600 ELLPLV LLVALVHLEHNEHELEHNEHLEHNEHLEHNEHLEHNEH 1667
 OY 354 LPABAA 359
 DB 1668 LSLEAS 1673

Search completed: April 2, 2003, 11:53:17
 Job time : 18 secs

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200


```

; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
; US 09 568 992 51

Query Match      16.7% Score 318.5, ID 4, Length 446
Best Local Similarity 22.7% Pctd. No. 3, Gap 24
Matches 82, Conservative 78, Mismatches 11, Indels 0, Gaps 11

CY 13 TWNSRVLDVWVSDTRQUMSIBKATINNAIKZALATVIFPLDIEVTRGSSLDIN 72
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 TWNTATASEMIFKSEVCHIDAP----FEVVALDIFZMMZEEKEKHIML 56
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 73 IFAPVPPFPAINTVDSMPYACQVQNTWMAWMIETETPOTMELVAPV 112
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 LKCKREKSAIKLLNEVCHLEHREELRESLAWSEFSENNKELT 115
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 133 GLSHQVBEPLNQSSEFFWINQF---LSLENSPQALHAY 199
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 GQVNTTETRETFVGLTFHNDKZELFKERLKEELKELWVPSQVWVNDLE 196
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 193 DSALQI 214
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 NSGNHEDMAGCTDQVLCQHLTGGILHETTFAPSK 214
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 232 FGVETL 276
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 ---EPMKALLQWVDPHPLPNSGIVSESTENLEKELFKVLYMHH 271
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 47 484VYLFHNSAGEMP 428
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 LLSMFLAPQVIGELLNGLHFFKAKQCOAHLEW 427
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 329 QI 330
Db 328 RL 329

RESULT 6
US 09 657 474 51
; Sequence 51, Application US/03659424
; Patent No. 6299462
; GENERAL INFORMATION:
; APPLICANT: Genet, H.
; APPLICANT: Fremmet, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSIS
; FILE REFERENCE: 7853 338
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: 09/234,114
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/166,064
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/088,112
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/078,044
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-657-474-51

Query Match      16.7% Score 318.5, ID 4, Length 446
Best Local Similarity 22.7% Pctd. No. 3, Gap 24
Matches 82, Conservative 78, Mismatches 11, Indels 0, Gaps 11
```

```

; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
; US 09 568 992 51

Query Match      16.7% Score 318.5, ID 4, Length 446
Best Local Similarity 22.7% Pctd. No. 3, Gap 24
Matches 82, Conservative 78, Mismatches 11, Indels 0, Gaps 11

CY 13 TWNSRVLDVWVSDTRQUMSIBKATINNAIKZALATVIFPLDIEVTRGSSLDIN 72
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 TWNTATASEMIFKSEVCHIDAP----FEVVALDIFZMMZEEKEKHIML 56
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 73 IFAPVPPFPAINTVDSMPYACQVQNTWMAWMIETETPOTMELVAPV 112
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 LKCKREKSAIKLLNEVCHLEHREELRESLAWSEFSENNKELT 115
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 133 GLSHQVBEPLNQSSEFFWINQF---LSLENSPQALHAY 199
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 GQVNTTETRETFVGLTFHNDKZELFKERLKEELKELWVPSQVWVNDLE 196
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 193 DSALQI 214
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 NSGNHEDMAGCTDQVLCQHLTGGILHETTFAPSK 214
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 232 FGVETL 276
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 ---EPMKALLQWVDPHPLPNSGIVSESTENLEKELFKVLYMHH 271
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 47 484VYLFHNSAGEMP 428
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 LLSMFLAPQVIGELLNGLHFFKAKQCOAHLEW 427
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 329 QI 330
Db 328 RL 329

RESULT 6
US 09 657 474 51
; Sequence 51, Application US/03659424
; Patent No. 6299462
; GENERAL INFORMATION:
; APPLICANT: Genet, H.
; APPLICANT: Fremmet, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSIS
; FILE REFERENCE: 7853 338
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: 09/234,114
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/166,064
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/088,112
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/078,044
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-657-474-51

Query Match      16.7% Score 318.5, ID 4, Length 446
Best Local Similarity 22.7% Pctd. No. 3, Gap 24
Matches 82, Conservative 78, Mismatches 11, Indels 0, Gaps 11
```

```

; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
; US 09 568 992 51

Query Match      16.7% Score 318.5, ID 4, Length 446
Best Local Similarity 22.7% Pctd. No. 3, Gap 24
Matches 82, Conservative 78, Mismatches 11, Indels 0, Gaps 11

CY 13 TWNSRVLDVWVSDTRQUMSIBKATINNAIKZALATVIFPLDIEVTRGSSLDIN 72
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 TWNTATASEMIFKSEVCHIDAP----FEVVALDIFZMMZEEKEKHIML 56
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 73 IFAPVPPFPAINTVDSMPYACQVQNTWMAWMIETETPOTMELVAPV 112
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 LKCKREKSAIKLLNEVCHLEHREELRESLAWSEFSENNKELT 115
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 133 GLSHQVBEPLNQSSEFFWINQF---LSLENSPQALHAY 199
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 GQVNTTETRETFVGLTFHNDKZELFKERLKEELKELWVPSQVWVNDLE 196
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 193 DSALQI 214
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 NSGNHEDMAGCTDQVLCQHLTGGILHETTFAPSK 214
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 232 FGVETL 276
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 ---EPMKALLQWVDPHPLPNSGIVSESTENLEKELFKVLYMHH 271
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 47 484VYLFHNSAGEMP 428
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 LLSMFLAPQVIGELLNGLHFFKAKQCOAHLEW 427
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
CY 329 QI 330
Db 328 RL 329

RESULT 6
US 09 657 474 51
; Sequence 51, Application US/03659424
; Patent No. 6299462
; GENERAL INFORMATION:
; APPLICANT: Genet, H.
; APPLICANT: Fremmet, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSIS
; FILE REFERENCE: 7853 338
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: 09/234,114
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/166,064
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/088,112
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/078,044
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 1.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-657-474-51

Query Match      16.7% Score 318.5, ID 4, Length 446
Best Local Similarity 22.7% Pctd. No. 3, Gap 24
Matches 82, Conservative 78, Mismatches 11, Indels 0, Gaps 11
```




Genome version: 1.0.1.1.4578
Date: 07/19/03
Author: Genentech Ltd.

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578
Human genome sw: 1.0.1.1.4578
Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Human genome sw: 1.0.1.1.4578

Accession	ID	Description
AA01344	AA01344	Human gene 1, encoded
AB01344	AB01344	Human gene 1, encoded
AC01344	AC01344	Human gene 1, encoded
AD01344	AD01344	Human gene 1, encoded
AE01344	AE01344	Human gene 1, encoded
AF01344	AF01344	Human gene 1, encoded
AG01344	AG01344	Human gene 1, encoded
AH01344	AH01344	Human gene 1, encoded
AI01344	AI01344	Human gene 1, encoded
AJ01344	AJ01344	Human gene 1, encoded
AK01344	AK01344	Human gene 1, encoded
AL01344	AL01344	Human gene 1, encoded
AM01344	AM01344	Human gene 1, encoded
AN01344	AN01344	Human gene 1, encoded
AO01344	AO01344	Human gene 1, encoded
AP01344	AP01344	Human gene 1, encoded
AQ01344	AQ01344	Human gene 1, encoded
AR01344	AR01344	Human gene 1, encoded
AS01344	AS01344	Human gene 1, encoded
AT01344	AT01344	Human gene 1, encoded
AW01344	AW01344	Human gene 1, encoded
AX01344	AX01344	Human gene 1, encoded
AY01344	AY01344	Human gene 1, encoded
AZ01344	AZ01344	Human gene 1, encoded

Genome version: 1.0.1.1.4578
Date: 07/19/03
Author: Genentech Ltd.

Human genome sw: 1.0.1.1.4578

11	628	32.9	148	21	AA01344	Human gene 1, encoded
12	604	31.6	139	22	AA01344	Human gene 1, encoded
13	604	31.6	139	23	AA01344	Human gene 1, encoded
14	604	31.6	139	24	AA01344	Human gene 1, encoded
15	604	31.6	139	25	AA01344	Human gene 1, encoded
16	579	30.3	145	26	AA01344	Human gene 1, encoded
17	494	25.9	112	27	AA01344	Human gene 1, encoded
18	494	25.9	112	28	AA01344	Human gene 1, encoded
19	489	25.6	112	29	AA01344	Human gene 1, encoded
20	489	25.6	112	30	AA01344	Human gene 1, encoded
21	474	24.8	114	31	AA01344	Human gene 1, encoded
22	433.5	22.7	110	32	AA01344	Human gene 1, encoded
23	433.5	22.7	110	33	AA01344	Human gene 1, encoded
24	355	18.6	143	34	AA01344	Human gene 1, encoded
25	336.5	17.6	130	35	AA01344	Human gene 1, encoded
26	321	16.8	145	36	AA01344	Human gene 1, encoded
27	320	16.8	145	37	AA01344	Human gene 1, encoded
28	320	16.8	145	38	AA01344	Human gene 1, encoded
29	320	16.8	145	39	AA01344	Human gene 1, encoded
30	320	16.8	145	40	AA01344	Human gene 1, encoded
31	320	16.8	145	41	AA01344	Human gene 1, encoded
32	319.5	16.7	145	42	AA01344	Human gene 1, encoded
33	319	16.7	145	43	AA01344	Human gene 1, encoded
34	318.5	16.7	145	44	AA01344	Human gene 1, encoded
35	307	16.1	145	45	AA01344	Human gene 1, encoded
36	300	15.7	145	46	AA01344	Human gene 1, encoded
37	300	15.7	145	47	AA01344	Human gene 1, encoded
38	292.5	15.3	145	48	AA01344	Human gene 1, encoded
39	292.5	15.3	145	49	AA01344	Human gene 1, encoded
40	292.5	15.3	145	50	AA01344	Human gene 1, encoded
41	292.5	15.3	145	51	AA01344	Human gene 1, encoded
42	292.5	15.3	145	52	AA01344	Human gene 1, encoded
43	292.5	15.3	145	53	AA01344	Human gene 1, encoded
44	292.5	15.3	145	54	AA01344	Human gene 1, encoded
45	292.5	15.3	145	55	AA01344	Human gene 1, encoded

ALIGNMENTS

RESULT 1

AA01344
ID: AA01344 Standard; Protein: 348 AA.

XX: AA01344

XX: 07 AUG 2001 (first entry)

XX: Human gene 1, encoded secreted protein (HKG1), 348 AA.

XX: Human; secreted protein; proliferative disorder; cancer; tumor;
XX: fetal abnormality; developmental abnormality; hepatopulmonary disorder;
XX: immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX: inflammation; allergy; neurological disorder; Alzheimer's disease;
XX: Parkinson's disease; cognitive disorder; schizophrenia; alcohol;
XX: skin disorder; psoriasis; sepsis; diabetes; alcoholism;
XX: cardiovascular disorder; angiodysplasia; kidney disorder;
XX: gastrointestinal disorder; pregnancy-related disorder; amenorrhea;
XX: endocrine disorder; infection; wound healing; vitellaria;
XX: cell culture; chemotaxis; food additive;
XX: binding partner identification.

XX: Homo sapiens

XX: Key

XX: Peptide

XX: Protein

XX: Note: "Mature secreted protein"

XX: W42010337.11

XX: 10-MAY 2001

XX: 10-MAY 2001



XX PF 17-OCT-2000; 2000W01529664
 XX PR 02-NOV-1999; 99NS-0164085
 XX PR 17-DEC-1999; 99NS-0172411
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
 PI Komatsoulis G, Poser CA, Fokien SM, Guan PC, Young PE, Floriano KA;
 PI Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
 XX WP1: 2001-328782/34.
 XX DR N-PSUB; AAD08192
 XX Novel human secreted proteins and nucleic acids for diagnosing,
 PT preventing and treating neurological, cardiovascular, infectious,
 PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
 PT ovarian cancer
 XX Claim 11: Page 387-390; 421pp; English.
 XX AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
 CC protein genes and AAK03764-AAK03786 represent the proteins they encode.
 CC AAK03787-AAK04800 represent human secreted protein fragments or variants.
 CC The genes and their secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 19 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumors, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angioecenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infectious. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands of binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention.
 XX SQ Sequence 766 AA;

Query Match 100.0%, Score 4081, Dn 22, Length 766;
 Best local Similarity 100.0%, proc No. 0;
 Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWRSPAGAELESLMALWENTAIETWVAVAAVSTQDIATSPFVLLSLKGPHPSPQFY 60
 DB 1 MWRSPAGAELESLMALWVWIAISLHGWVAVAAVSTQDIATSPFVLLSLKGPHPSPQFY 60
 QY 61 TDVPSRPGSTFYKYIYEPFQWVNNI AVEPPNPLASPLAPFPFNPIPLPPTTL 120
 DB 61 TDVPSRPGSTFYKYIYEPFQWVNNI AVEPPNPLASPLAPFPFNPIPLPPTTL 120
 QY 121 GGTENLIRKRYGHHPLZADIDPSSIIIPVWVFFLSPPAPGSSSTNSSVPTFHQ 180
 DB 121 GGTENLIRKRYGHHPLZADIDPSSIIIPVWVFFLSPPAPGSSSTNSSVPTFHQ 180
 QY 181 AASYFIDRSTLPPHHIQTASTATKVTFTPTGPGGTSNYNLSGSVSVLYVSPENKQL 240
 DB 181 AASYFIDRSTLPPHHIQTASTATKVTFTPTGPGGTSNYNLSGSVSVLYVSPENKQL 240

QY 241 QGLOVLEPVTQPRFVQAAI SYTACNSPPTTCKNINWTHCZIEFFPCTCSMLQAME 300
 DB 241 QGLOVLEPVTQPRFVQAAI SYTACNSPPTTCKNINWTHCZIEFFPCTCSMLQAME 300
 QY 301 PNIIRITETWKAAYNSDPFSPKPKI PMNVEI NSTIMHIWIMDSNPPPPYQLEN 360
 DB 301 PNIIRITETWKAAYNSDPFSPKPKI PMNVEI NSTIMHIWIMDSNPPPPYQLEN 360
 QY 361 SNKQLFLKAKIVKULSLSKSKCHKOPLISLPPTSTYWLTPQSELYCNEWLLSEFS 420
 DB 361 SNKQLFLKAKIVKULSLSKSKCHKOPLISLPPTSTYWLTPQSELYCNEWLLSEFS 420
 QY 421 PPTHSTQTPNQVVAAPFPCIVTQVTPASACITCAINPRTNPTQVYMI SQCLCKPEVAE 480
 DB 421 BEHHSITPNIQVVAAPFPCIVTQVTPASACITCAINPRTNPTQVYMI SQCLCKPEVAE 480
 QY 481 STDRYIGFTDQLEMKYLLQKTDKRIEVHAFISNIMKINSWEDISWKKRMILLKSN 540
 DB 481 STDRYIGFTDQLEMKYLLQKTDKRIEVHAFISNIMKINSWEDISWKKRMILLKSN 540
 QY 541 KYKSSLVHMLGLSLQICITKNTSLPEVLAVVNPFGSSHSFMPVNFNSPDPWERTK 600
 DB 541 KYKSSLVHMLGLSLQICITKNTSLPEVLAVVNPFGSSHSFMPVNFNSPDPWERTK 600
 QY 601 LQPLQCTNWTLLGNKKWTFEFTVHYVRSKPSNPNPNPNPNPNPNPNPNPNPNPN 660
 DB 601 LQPLQCTNWTLLGNKKWTFEFTVHYVRSKPSNPNPNPNPNPNPNPNPNPNPNPN 660
 QY 661 MKINNTQVFGYSMMHDPKAIHLLIQLQVYVFGSGSALQILEHRYNKLISPCQRR 720
 DB 661 MKINNTQVFGYSMMHDPKAIHLLIQLQVYVFGSGSALQILEHRYNKLISPCQRR 720
 QY 721 LDFSCLLPHLPKLTSEVVRISQALQAFNAKLENTMDYETIKLS 766
 DB 721 LDFSCLLPHLPKLTSEVVRISQALQAFNAKLENTMDYETIKLS 766

RESULT 2
 AAG65480
 ID AAG65480 standard: Protein; 766 AA.
 AC AAG65480;
 XX 27-AUG-2002 (first entry)
 DE Human albumin fusion protein #2155.
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA, cancer, reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antiinfertility; antiinflammatory; antiulcer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic
 XX Homo sapiens.
 OS Synthetic.
 XX WQ200177137-A1.
 XX 18-OCT-2001.
 XX 12-APR-2001; 2001W0-US11988.
 XX 12-APR-2001; 2000US-254358P.
 XX 25-APR-2001; 2000US-199384P.
 XX 21-DEC-2001; 2000US-256931P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Poser CA, Haseltine WA;
 PI

hyperproliferative disorders, sleep pattern, cardiovascular disorders;
reproductive disorder; digestive system disorders; behavioural disorder;
Homo sapiens.

W0200216390 AI.

28 FEB-2002.

17 JAN-2001; 2001WO-US01435.

18 AUG-2000; 2000US-226382P.

(HUMA) HUMAN GENOME SCI INC.

Rosen CA, Komatsoulis CA, Baker KP, Hirse CH, Soppet DR, Olson HS;
Moore PA, Wei P, Ebner K, Duan DR, Shi Y, Choi GH, Piscella M;
NI J;

WPI; 2002-304113/34.

An isolated nucleic acid molecule (1) comprising a polypeptide which
encodes a polypeptide useful in the diagnosis and treatment of
disorders e.g. immune disorders -

Disclosure: Page 59; 534pp; English.

AAU33692-AAU33736 represent cDNAs corresponding to 21 human secreted
protein genes, and AAU21191-AAU21225 represent the proteins they encode.
AAU21236-AAU21280 represent human secreted protein fragments. The genes
and their corresponding secreted proteins are useful for preventing,
treating or ameliorating medical conditions, e.g., by protein or gene
therapy. Pathological conditions can be diagnosed by determining the
amount of the new protein in a sample or by determining the presence of
mutations in the new genes. Specific uses are described for each of the
21 genes, based on the tissues in which they are most highly expressed,
and include developing products for the diagnosis or treatment of
immune or autoimmune diseases e.g. AIDS (acquired immune deficiency
syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
and breast cancer, neurological diseases e.g. Alzheimer's disease,
Parkinson's disease, Huntington's disease, Tourette syndrome,
meningitis, demyelinating disease, peripheral neuropathies, toxic
trauma, congenital malformations, spinal cord injuries, multiple
neuropathies induced by neurotoxins, peripheral neuropathies, multiple
sclerosis, ischaemia and infarction, haemorrhages, schizophrenia, mania,
dementia, depression, panic disorder, learning disabilities, AIDS,
altered behaviours e.g. disorders in feeding, sleep patterns, balance
and perception, encephalitis, disorders in cardiovascular, neural/
sensory, reproductive and digestive systems, behavioural disorders and
hyperproliferative disorder. The present sequence represents human
secreted protein fragment referred to in the disclosure of the invention.

Sequence 347 AA:

Query Match 41.4%; Score 1691; Ds 24; Length 337;
Best local Similarity 100.0%; Pred. No. 1,26-139;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MIWRSAGAEHLSMALWVWIALSLHCVLAAVAASVQHAISPEWILSDKGPFRSDEY 60

1 MIWRSAGAEHLSMALWVWIALSLHCVLAAVAASVQHAISPEWILSDKGPFRSDEY 60

61 TDVDSRQGFSTRYKLYEFGKRWKNNIATVPRNPLGSGPLAEPFRNPLGCPDPPT 120

61 TDVDSRQGFSTRYKLYEFGKRWKNNIATVPRNPLGSGPLAEPFRNPLGCPDPPT 120

121 QQTTFNIKKYGTHTFLSATLGGFFSITFVDFKFLSKPAFGSDSTNNSSVTLETLHQL 180

121 QQTTFNIKKYGTHTFLSATLGGFFSITFVDFKFLSKPAFGSDSTNNSSVTLETLHQL 180

181 AASYFIDRUSTLPRLLHIOASTAIKVTETRTGPGCCSNYNLDSSVSVIYOSPENKIQ 240

181 AASYFIDRUSTLPRLLHIOASTAIKVTETRTGPGCCSNYNLDSSVSVIYOSPENKIQ 240

QY 241 QSLQVLLFTYLYFFEVQAALSYIAINSEETLREKSWHLSASLILGTHLMGLVAML 300
DB 241 QSLQVLLFTYLYFFEVQAALSYIAINSEETLREKSWHLSASLILGTHLMGLVAML 300

QY 301 ENLRITETWKAYNSDFEES 320

DB 301 ENLRITETWKAYNSDFEES 320

RESULT 7

AAM58347

ID AAM58347 standard; Protein; 478 AA.

XX AAM58347;

XX 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 40462.

KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer.

XX Homo sapiens.

XX WC200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US 0180312.

XX 26-MAY-2000; 2000US 0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 04-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US 0234687.

XX 27-SEP-2000; 2000US 0236359.

XX 04-OCT-2000; 2000US-0034263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains.

XX Example 4; SEQ ID NO: 30452; 650pp - Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.

XX Sequence 378 AA:

Query Match 38.0%; Score 1550; Ds 22; Length 378;
Best local Similarity 74.7%; Pred. No. 1,5-127;

Matches 278; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 395 RTSTWLTWTLTQFLYCNENGLISFSETHSCTPNQVQVCTATPLPTVSDASA-ITAP 474

DB 7 RSLSWNNRIQSLLYCGESTFEGTELESHSCTPYDQSSQAFIPCALGEGPAWACAP 60

QY 455 DNRTRCGTCNTGYMLSGGLCKEYVAESTDHYTGETHQGLEMAKYLQKIDRRLEVHAIF 514

DB 67 DNSTRTGSGNPGYVLAAQLOPPEVAESLENLEGLTELEQLQLELYLQKQDSKEVHSIF 126

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234487.
 PR 27-SEP-2000; 2000US-0234459.
 PR 04-OCT-2000; 2000GB-0624263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PA Penn SG, Hanzel DR, Chen W, Rank DR;

XX WPI; 2001 483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
 brains -

XX Example 4: SEQ ID NO: 30424; 660pp - Sequence Listing: English.

XX The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 brain. They can be used to measure gene expression in brain cell samples,
 which may enable the diagnosis and improved treatment of nervous system
 diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 epilepsy and cancers. The present sequence is a protein encoded by one of
 the probes of the invention.

XX Sequence 370 AA:

Query Match 26 280 Score 1097.6; FR 23; Length 470;

Best Local Similarity 60 180 Prod No 2 70-91;

Matches 10; Conserved 10; Mismatch 10; Index 10; Gap 5;

QY 396 LSYWLTIRUSLYCNGINGLDSFSEHDSCHCPNDVVNTAFCTVDSACITDAPD 455

Db 1 TIOGILARVUSLYCNGENSPWTFETESQSVVHSTTHQPPPEVTSNNSPAMSLA 66

QY 456 NPTQITNTSYMLSCGKKEFAVES-THYVAGEHID--GQEMKYLKLIKPKTEVH 511

Db 61 NISGSCNKGYLEYRPPPPNVNVSPPSQFSEFDIDDEPKYKIDKMSPPGVH 120

QY 512 AIFSNMPINSPDSWPFMLITTPSNFYSSIVMTGFGIQTLPSTSTFFVLAV 571

Db 121 TTFISNEITDTEFPWFPMSTIFKSNKMPFIHMVGNMSPICMPSNLSIFEEV 180

QY 572 VVNPEZSSHSFWMVNVNSLAWKATLQVQVYNNWLLTNNKPLFPFVVPVPS 631

Db 181 VYNPFSHSFWMVNPFFGFYFWEKTRIFQNSQYNNWLLTNNKPLFPFVVPVPS 740

QY 632 EIKSNKPNKNSTVYFPEFIFASPLANVKKINNKAVSYSMFPGEPALPDLGLVY 691

Db 240 KTRIPPIENFEGAGQVDVDSKSKQPVKIKISWGVKYSIPKNADIPSAVGVNQS 798

QY 692 TQNSQ----DSALLQLFIDPPVNTISP---PQGPPIEFSTLPPIRIFTSTFVVPVPS 744

Db 290 TQNSQFYSSSSVMLETFEPTNELAPPVAPKPKQLDPSMKIKPLKLNSELIPVNI 808

QY 745 ALQAFNAKIDNMDVDTKE 765

Db 359 ALDIYNTLTKSQSTAKIG 474

RESULT 10

ABH14717

Id ABH14717 standard; Prototag; 121 AA

XX ABB14717;

XX 23 JAN 2002 (first entry)

XX Human nervous system related polypeptide Seq ID No 3374

XX Human; neurotropic; neuroproliferative; cytostatic; dermatological; virucide;

XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;

XX antiparkinsonian; antiskink; antianemic; antarthritic; cancer;

XX antirheumatic; hepatotropic; corobroprotective; antioflammatory;

KW anti-allergic; antidiabetic; antiulcer; anticonvulsant; antitumoral;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccines;
 XX Homo sapiens.

XX WC200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065

XX 04-FEB-2000; 2000US-0180628

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186550.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 26-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220763.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0225279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226686.

XX 24-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0229513.

XX 06-SEP-2000; 2000US-0230437.

XX 06-SEP-2000; 2000US-0230438.

XX 08-SEP-2000; 2000US-0231242.

XX 08-SEP-2000; 2000US-0231243.

XX 08-SEP-2000; 2000US-0231244.

XX 08-SEP-2000; 2000US-0231413.

XX 08-SEP-2000; 2000US-0231414.

XX 08-SEP-2000; 2000US-0232080.

XX 08-SEP-2000; 2000US-0232081.

XX 12-SEP-2000; 2000US-0231968.

XX 14-SEP-2000; 2000US-0232397.

XX 14-SEP-2000; 2000US-0232398.

XX 14-SEP-2000; 2000US-0232399.

XX 14-SEP-2000; 2000US-0232400.

XX 14-SEP-2000; 2000US-0232401.

XX 14-SEP-2000; 2000US-0233063.

XX 14-SEP-2000; 2000US-0233064.

XX 14-SEP-2000; 2000US-0233065.

XX 21-SEP-2000; 2000US-0234223.

CC mutations in the new genes. Specific uses are described for each of the
 CC 21 genes based on the tissues in which they are most highly expressed.
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases e.g. AIDS (acquired immune deficiency
 CC syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
 CC and breast cancer, neurological diseases e.g. Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease, Lou Gehrig syndrome,
 CC meningitis, demyelinating disease, peripheral neuropathies, neoplasia,
 CC trauma, congenital malformations, spinal cord injuries, toxic
 CC neuropathies induced by neurotoxins, peripheral neuropathies, multiple
 CC sclerosis, ischaemia and infarction, haemorrhages, schizophrenia, mania,
 CC dementia, depression, panic disorder, learning disabilities, AIDS,
 CC altered behaviours e.g. disorders in feeding, sleep patterns, balance
 CC and perception, encephalitis, disorders in cardiovascular, neural/
 CC sensory, reproductive and digestive systems, behavioural disorders and
 CC hyperproliferative disorder. The present sequence represents human
 CC secreted protein fragment referred to in the disclosure of the invention.

XX Sequence 73 AA:

Query Match 9.7%; Score 396; DB 24; Length 73;
 Best Local Similarity 100.0%; Pred. No. 5,56-27;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIMSRAGAEFLSMLALWELIALSLHCWVLAVALSDHATSPDFWLLSDKGPFRHSQFY 60
 Db 1 MIMSRAGAEFLSMLALWELIALSLHCWVLAVALSDHATSPDFWLLSDKGPFRHSQFY 60

QY 61 TDFVDRSRKGFST 73

Db 61 TDFVDRSRKGFST 73

RESULT 15

ABB42190

ID ABB42190 standard; Peptide; 78 AA.

XX ABB42190;

XX 01-FEB-2002 (first entry)

XX Peptide #4841 encoded by breast cell single exon nucleic acid probe.

XX Human: microarray; single exon probe; gene expression; breast;

XX disease; cancer.

XX Homo sapiens

XX W0200157271.A2.

XX 09-AUG 2001.

XX 30-JAN-2001; 2001W0-0500562.

XX 04-FEB 2000; 2000US-0180312.

XX 24-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0242563.

XX (MOLEC.) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001 496933/54.

XX New spatially addressable set of single exon nucleic acid probes,

XX useful for measuring gene expression in sample derived from human

XX breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 15158; 327bp; sequence listing; English.

XX The invention relates to a spatially addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or diagnosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from Wipo at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 78 AA:

Query Match 8.0%; Score 327; DB 22; Length 78;
 Best Local Similarity 70.5%; Pred. No. 76-21;
 Matches 55; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 243 LOVLLPDYLOEPFVQALSYATCNSESEFICKENQWICGPKIPFICWTPSKMDQAMEEN 602

Db 1 LOVLLPDYLERFVAAALSYITCSSHGRLVCKENQWICGPKIPFICWTPSKMDQAMEEN 60

QY 303 LKRITETWKAYNSDFEES 320

Db 61 LIQIQDSWATHNRQPEES 78

Search completed: May 19, 2003, 10:31:18

Job time : 83 secs

Genome version: 1.1.1 (15_457A)
Date: 10/29/03

Search used sw model:

10/29/03 10:29:53 Search for 30 seconds
(with 10 alignments)
10/29/03 10:29:53 Search for 30 seconds
(with 10 alignments)

SEARCH RESULTS SUMMARY

902

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

10/29/03 10:29:53

number of results predicted by chance to have a
score equal to the score of the result being printed,
7 analysis of the top 10 scores distribution.

SUMMARY

Seq ID	DB	ID	Description
100	4	US-09-627-376-7	Sequence 7, Appl
101	4	US-09-627-376-7	Sequence 22, Appl
102	4	US-09-627-376-7	Sequence 55, Appl
103	4	US-09-627-376-7	Sequence 22, Appl
104	4	US-09-627-376-7	Sequence 55, Appl
105	4	US-09-627-376-7	Sequence 2, Appl
106	4	US-09-627-376-7	Sequence 4, Appl
107	4	US-09-627-376-7	Sequence 4, Appl
108	4	US-09-627-376-7	Sequence 125, Appl
109	4	US-09-627-376-7	Sequence 7, Appl
110	4	US-09-627-376-7	Sequence 2, Appl
111	4	US-09-627-376-7	Sequence 2, Appl
112	4	US-09-627-376-7	Sequence 84, Appl
113	4	US-09-627-376-7	Sequence 84, Appl
114	4	US-09-627-376-7	Sequence 86, Appl
115	4	US-09-627-376-7	Sequence 86, Appl
116	4	US-09-627-376-7	Sequence 90, Appl
117	4	US-09-627-376-7	Sequence 90, Appl
118	4	US-09-627-376-7	Sequence 131, Appl
119	4	US-09-627-376-7	Sequence 131, Appl
120	4	US-09-627-376-7	Sequence 17, Appl
121	4	US-09-627-376-7	Sequence 17, Appl
122	4	US-09-627-376-7	Sequence 140, Appl
123	4	US-09-627-376-7	Sequence 140, Appl
124	4	US-09-627-376-7	Sequence 15, Appl
125	4	US-09-627-376-7	Sequence 15, Appl
126	4	US-09-627-376-7	Sequence 28, Appl
127	4	US-09-627-376-7	Sequence 28, Appl
128	4	US-09-627-376-7	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-09-627-376-7

Sequence 7, Application US/09627376

Patent No. 6342385

GENERAL INFORMATION:

APPLICANT: Q1, Fengxia

TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS

FILE REFERENCE: 174,622-2

CURRENT APPLICATION NUMBER: US/09627376

CURRENT PUBLICATION NUMBER: 174,622-2

NUMBER OF SEQ. IN NO. 174,622-2

SEQ. ID NO. 7

LENGTH: 990

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-627-376-7

Query Match

Best Local Similarity 18.9%; Pred. No. 0.1;

Matches 164; Conservative 125; Mismatches 275; Indels 245; Gaps 42;

QY	119	LCQVLEHLLKKYQ-----HPLLSALHGHSHSLPLVHAKLSKAKESNSTINS	169
DB	139	TLKVVANNALEKSNDFWLLTRSEFGLMNSPDRILEITY-----ESNLIIDVINE	190
QY	170	SS--VTLETHQLAASYFISFSLTFLHRIHQTASTAK-----ATETETETELPKNVLE-	222
DB	191	TEEPISYVHLLDIAEKFSOSSDVK-----PYQLHLLKEPLLEKESLIDP-DNP	241
QY	223	LDSVSSVLVSPENKIGLOGL---QVLLPYLQEPVVAALSYA-NREFETICKEN	276
DB	242	LQWFINILETPQNSLELFTETAYAMLYTENTGEMHNSLALEREMFIVKANAYL	401
QY	277	--DCWCACGKPKFCPCPSMULGAMFNILPILFELW-KATNSHRESHDEKLPMPK	429
DB	302	PVLDYDAELPLAGHTKSSST-----QNLKVLSTSSAVNNSVETKRYDEFLIAYVE	455
QY	330	--LPMNPLNLTSTHMLWTMNSFPPPYALE-NSMFLFLKALGVHPL	483
DB	356	QVHPLQLEHNS-----TSGLGPKGYSQTEVSKNNHLSKNKLTEDLQKFKALRD	408
QY	384	HKOPLILSPQRTSTY-WLTRIQSLEYCNEN-----GLLSFSEETHSTCT	428
DB	409	GKPIILNHLHFKLNEHPLWILSEELYLYNFKFVPLNSLVSQMLNIPHIS	467
QY	429	PNHGVGCTATPQIVGLASACILCAHNRKGGKNDY--MNSDQKPEVAHS	481
DB	468	FN-----TIVKNNVKKIKETEAYPNIIIVLNEVFFVIGKQNMIMNSL-KNQLLEL	520

QY 482 TDHYHGFPTDGLD-----FMKVIHKTDEPIEVAIFISNDMPNLNSWDSWPKKML 534
 Db 521 RNYTKKEMSIINDIYVPAISEHLYFYSKKYEKV---IPVNNM--PNYNGSKLRF 574
 QY 535 LTKSNKYSKSIYHMIHGL-----SIQICHTK 561
 Db 575 LEVNSDFQ-NITPITGLSLSYHNPALYYKDIIKIPETWIRKSEAKTILDSLKNWLT- 632
 QY 562 NSTLEIVL---AVYVNFPGSHSGWPMVNVNSP-----PWERTKLEL 603
 Db 633 NNNVPEVRKKYTYQIYIYISPDITWLPQSIKHSPIOLLNVHVSCTNTEILELV 692
 QY 604 PL-----QCYNWLT-----GNKKWTFEIVHI-----YLR- 636
 Db 693 PETHSWAHQIYHQAQNIYTLFDSGSKIKYKPYAKIYVNNKOPQTSFLOKEYPHLLIKYIKL 752
 QY 631 -----SRIKNGUNGNE-----SIYEPHLEFDPSKNIQYKNNIQVFC 670
 Db 753 PENIQWYIYIKODKGRDSTELPIYVYEDKQVGLYSRPIFWATKAP-----KNIQISG 805
 QY 671 YSM-HHDEPAIR-----DLIIQIYIYPTQCSQDSALIQLEIHDR 709
 Db 806 YEISEYIPESARYGKNGKYSIHSFYYIUSIILDLIQ-----KKKQTIKRTS 854
 QY 710 VN-----KLSPPQRL--DLFSCLLRHRLK 734
 Db 855 LSTIRMLMKLSIQAQKLIKIKNIFAG--KKKIK 886

RESULT 2

US-09-183-861-22
 ? Sequence 22, Application 35/09183861
 ? Patent No. 6365165
 ? GENERAL INFORMATION:
 ? APPLICANT: Reed, Steven G.
 ? APPLICANT: Campos-Neto, Antonio
 ? APPLICANT: Webb, John R.
 ? APPLICANT: Dillon, David C.
 ? APPLICANT: Skeiky, Yasir A.W.
 ? TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
 ? NUMBER OF SEQUENCES: 87
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: SEED and BERRY LLP
 ? STREET: 6300 Columbia Center, 701 Fifth Avenue
 ? CITY: Seattle
 ? STATE: Washington
 ? COUNTRY: USA
 ? ZIP: 98104-7092
 ? COMPUTER READABLE FORM:
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? FILING DATE: US/09/183,861
 ? CLASSIFICATION:
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 09/022,765
 ? FILING DATE: 12-FEB-1998
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Maki, David J.
 ? REGISTRATION NUMBER: 31,392
 ? REFERENCE/DOCKET NUMBER: 210121.42003
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (206) 622-4900
 ? TELEFAX: (206) 682-6031
 ? INFORMATION FOR SEQ ID NO: 22:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 320 amino acids
 ? TYPE: amino acid
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? FRAGMENT TYPE: internal
 ? US-09-183-861-22

US-09-183-861-22
 Query Match 2.8%; Score 114.5; DR 4; Length 320;
 Best local Similarity 32.3%; Pred. No. 0.02;
 Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 4;
 QY 383 GHKQPLISLPQRTSTYWLRTQSFLYNENGLGSPFSETHISCTQPNQVVTAFLEPT 442
 Db 176 FRLSIACSVPNKKKGTGHSPL-----GADQGYSLSAIAISSSPTG----- 224
 QY 443 VRIASACITAFNPRTPIGNTNGYMLSGLEK 475
 Db 224 V---EHCNTGVNJDSTPCAYCNGTYVVSDEKCK 253
 RESULT 3
 US-09-183-861-55
 ? Sequence 55, Application 05/09183861
 ? Patent No. 6365165
 ? GENERAL INFORMATION:
 ? APPLICANT: Reed, Steven G.
 ? APPLICANT: Campos-Neto, Antonio
 ? APPLICANT: Webb, John R.
 ? APPLICANT: Dillon, David C.
 ? APPLICANT: Skeiky, Yasir A.W.
 ? TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
 ? NUMBER OF SEQUENCES: 87
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: SEED and BERRY LLP
 ? STREET: 6300 Columbia Center, 701 Fifth Avenue
 ? CITY: Seattle
 ? STATE: Washington
 ? COUNTRY: USA
 ? ZIP: 98104-7092
 ? COMPUTER READABLE FORM:
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? FILING DATE: US/09/183,861
 ? CLASSIFICATION:
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 09/022,765
 ? FILING DATE: 12-FEB-1998
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Maki, David J.
 ? REGISTRATION NUMBER: 31,392
 ? REFERENCE/DOCKET NUMBER: 210121.42003
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (206) 622-4900
 ? TELEFAX: (206) 682-6031
 ? INFORMATION FOR SEQ ID NO: 55:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 320 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS:
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? FRAGMENT TYPE: internal
 ? US-09-183-861-55

Query Match 2.8%; Score 114.5; DR 4; Length 320;
 Best local Similarity 32.3%; Pred. No. 0.02;
 Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 4;
 QY 383 GHKQPLISLPQRTSTYWLRTQSFLYNENGLGSPFSETHISCTQPNQVVTAFLEPT 442
 Db 176 FRLSIACSVPNKKKGTGHSPL-----GADQGYSLSAIAISSSPTG----- 224
 QY 443 VRIASACITAFNPRTPIGNTNGYMLSGLEK 475
 Db 224 V---EHCNTGVNJDSTPCAYCNGTYVVSDEKCK 253


```

QY 125 ENLTKRYCTHPTLSATLQOCESTLTPVSKPKLSKPAEGSDSTINSSSVTLTTLHQLAASY 184
DB 1239 RYEFKATDYPMR - QKAWMLILANE-HAKN-ESFCATTSAMVFKS - DTRF 1288
QY 185 FUDNSTPPLHHTQIATPAIKVTPTPDQCSNYE:NLDSVSVLVSPENKIQLOQLQ 244
DB 1289 AVTRDLTKNGVYQFKEN - IOCTS-OFSSITAVLLQYSHD 1327
QY 245 VILPDYQBRFVQALSYIA-NSGEFTI-KEND-WB-GPAPPE-NTPSMULQAMPENIL 304
DB 1328 - - - - -ACMSW - - - - -FLKEG CF - - - - -PASAAGKCEGNSR 1353
QY 305 RITEWKAYNSDFRSDERKLFMKKPLMNYFINTSTIMELTMSDNFQRYEQLNSMKQ 364
DB 1354 ELSEPTVYTGDEWTRITIAIPR - - - - -SLASSKTRFRLQESSQKNNPP - - - - - 1401
QY 365 LFLKAOKIVIKLPSKSKCHKQFQI:SLIPKRTISYVILIR:USFYCNKNGLLGSFSEETH 424
DB 1402 - - - - -FCLJCVYISIPCHS - - - - -YCSHCG - - - - - 1421
QY 425 SCTCPNDQVCTAFLPCTWQASA - - - - -CLTCAPNRP - - - - -T 458
DB 1422 - - - - -DCISQVCF - - - - -CULGTYAAGCTCVSNIPNHSMF-DREFGKLSPIWYKITGGQVGT 1473
QY 459 RCTONTGYML-SOGICKPE - - - - -VAFSTD - - - - -HYIGFEIDIAQOLE 496
DB 1474 GCOTLNDGSLVFNGLGKREARTVPLDTRNLSLVQFYIQLGSKTSGIYIITPRARYESLV 1533
QY 497 MKY - - - - -LLOKTRIRIKVHALFISNDMRINS - - - - -WPDPSWKRMI - - - - - 534
DB 1534 VOYSNDNGILWHFIRLPHDMSPTFPQITISIDIPRAKTPATAPFWQPOQCKHSAOWALG 1593
QY 535 - - - - -LTLKSNKYKSSLVIMILGSLQI-CLTKNSTLEPVLAVY 572
DB 1594 LVLGVNWSQDQPKLIESLILQANNYP - - - - -LQKQVVDLQJLSMITAL - - - - -IP 1642
QY 573 VNPFGG-SHSESMPFVNKNS-PUMKPKLILPLOC - - - - -YNNWILILGNKWK 620
DB 1643 TENIGNPRYAFATGDFIVSESELQWE - - - - -MNM-K-SKPE-SHAR-IQLQYSLNN-KIWWL 1697
QY 621 FFE - - - - -TVH - - - - -LYUKRSIKSGPNGNHS:VYRPLRFDPSRNLGYMKINNIQ 667
DB 1698 VTRCEVPTIGCVHYTESSTYISRRQ - - - - -NWRVTVYVILATNSPTEPRFWIQTN - - - - - 1750
QY 668 VFGYSMIPOEALPILITQLVPPY 691
DB 1751 - - - - -YTVGADSWAIDNVILASGCPW 1771

```

RESULT 7

```

US-09-541-782-4
: Sequence 4, Application US/09541782
: Patent No. 6284480
: GENERAL INFORMATION:
: APPLICANT: Nislow, Corey
: APPLICANT: Sakowicz, Roman
: APPLICANT: Berand, Christophe
: TITLE OF INVENTION: Antifungal Assay
: FILE REFERENCE: 1015
: CURRENT APPLICATION NUMBER: US/09/541.782
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 1038
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-541-782-4

```

```

Query Match 2.5%; Score 103.5; DB 4; Length 1038.
Best local Similarity 20.5%; Pred. No. 1.6;
Matches 83; Conservative 76; Mismatches 167; Indels 79; Gaps 20;

```

```

QY 39 HATSPFDWL LSKGKPFHR - - - - -SQEYTFDVPSPQGS- - - - -NLAHYR - - - - -NLI 89
DB 363 HITNAMEGLNLQKGLKUPQVASTKMNDFSSPSHTITITILYKHAELEFISKMNVL 422
QY 90 A - - - - -VERNPFQSPDLAPDFFRNIRLQKRNPTLQOITNENLKKYCTHR - - - - -L 136
DB 423 AGSENIINSGALNORAKIAGSINOSLILGR - - - - -VINALVUKSG-HIFFPESKILRI 475
QY 137 LSATLGGESLTIFFVDRKLSKRAEGSDSTINSSSVTLTTLHQLAASYFIDRSTPLRH 196
DB 476 LQDSUGGNTKTAL-IATISPAKVTSHETCTSLLEYASKAKNINKKIQLOSGPMKIDLVKNI 534
QY 197 HIQIASTAIKTETRTGP - - - - -LGCSNYDNLDS-VSSVLVQSPENKIQLOQL - - - - -OVLIP 248
DB 535 TMEIAKISDLSLTKSKRTIYMSQHYKNLSGLESYKNEVQEKPEIESLISKNALLVK 594
QY 249 DYLOQRFVQAALSYIACNSGEFTICKKNKWC - - - - -HCGPKPFRONCFPSMIOA 298
DB 595 DKLKSK - - - - -ETIQSN - - - - -COTRESUKTITIDHLRAQIDKQHKTEFLISD 636
QY 299 MEENLRITETWKAYNSDFEE - - - - -SDEPKLF - - - - -MKRLPMNYFINTSTIMELWTK-D 448
DB 637 FNNKQKKTFFVQMAHUYKKPELILNCKPEMHITKEIKKIKSTLPTQINTMOORSTIQE 696
QY 349 SNFORRYEQLNSMKQLFKKAKIVHKLFSSKSKCKKQPLISLPR 393
DB 697 TNLQPNLLMIRNEVTLTMTWJKAELMY - - - - -KDCVKKILNESPK 738

```

RESULT 8

```

US-09-723-820-4
: Sequence 4, Application US/09723820
: Patent No. 6468760
: GENERAL INFORMATION:
: APPLICANT: Nislow, Corey
: APPLICANT: Sakowicz, Roman
: APPLICANT: Berand, Christophe
: TITLE OF INVENTION: Antifungal Assay
: FILE REFERENCE: 1015
: CURRENT APPLICATION NUMBER: US/09/723.820
: PRIOR FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: 09/541.782
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 1038
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-723-820-4

```

```

Query Match 2.5%; Score 103.5; DB 4; Length 1038.
Best local Similarity 20.5%; Pred. No. 1.6;
Matches 83; Conservative 76; Mismatches 167; Indels 79; Gaps 20;

```

```

QY 39 HATSPFDWL LSKGKPFHR - - - - -SQEYTFDVPSPQGS-TPYKIYR-ESPWKN-NLI 89
DB 363 HITNAMEGLNLQKGLKUPQVASTKMNDFSSPSHTITITILYKHAELEFISKMNVL 422
QY 90 A - - - - -VERNPFQSPDLAPDFFRNIRLQKRNPTLQOITNENLKKYCTHR - - - - -L 136
DB 423 AGSENIINSGALNORAKIAGSINOSLILGR - - - - -VINALVUKSG-HIFFPESKILRI 475
QY 137 LSATLGGESLTIFFVDRKLSKRAEGSDSTINSSSVTLTTLHQLAASYFIDRSTPLRH 196
DB 476 LQDSUGGNTKTAL-IATISPAKVTSHETCTSLLEYASKAKNINKKIQLOSGPMKIDLVKNI 534
QY 197 HIQIASTAIKTETRTGP - - - - -LGCSNYDNLDS-VSSVLVQSPENKIQLOQL - - - - -OVLIP 248
DB 535 TMEIAKISDLSLTKSKRTIYMSQHYKNLSGLESYKNEVQEKPEIESLISKNALLVK 594
QY 249 DYLOQRFVQAALSYIACNSGEFTICKKNKWC - - - - -HCGPKPFRONCFPSMIOA 298

```



```

Query Match          2.5%: Score 101.5; DB 5; Length 991,
Best Local Similarity 19.3%; Pred. No. 2.3; Idels 183; Gaps 29;
Matches 119; Conservative 76; Mismatches 183; Idels 187; Gaps 31;

QY 169 SSSVTLFTHQLAAS YFIDWSTLEPLHLHQ-----IASTAIKVIETFTGPL----- 215
DQ 86 SASITLQVLDVAICNISCWVKKHSSINCOPHFDLQNHGVSMVILKMTFTQAGHYLLF 145
QY 216 --GCSYNDNLDSVS-----SVLVQSPENKIQLOGLVLLPYLOERFVQALSYIAGNS 267
DQ 146 QSEATNYTILFTSIPNTLLTYLTPPYFKMENDALAVTISVDEPIVE-----VWLVD 201
QY 269 PGPCKPKP-----NIGWCHGCKPKP-----PNCNCPNM 294
DQ 202 QGE SKVEEFVAVVKKKEKVIHEIFDTLFCGAFNELLPEKTELTIDLTGTFQTLE 260
QY 295 DQAMEENLLPTE-----TWKAYNSDEESDEFKMKPLPMNYFNTSTIMHLW 345
DQ 261 FLKVGCEPLWIRKAVHVNHFQGLTWLEKNALEKGNFYEM-----SYSTNRTMIRLLF 314
QY 346 TMSNFOR--YEOLENSMKOIFIKAKKIVHKIP-----SISKPKYHKQPLI 389
DQ 315 AFVSSVARNDTGYTSSSKHPSQSALVTIVEKGFNATNSSEYIDQYEEFESVREK 374
QY 390 SLPKORTISYMLIRIQTSPYCNHCL-----LQSPSETHS-----CICPDQVVCYAF 438
DQ 375 AYPQIRCT--WTESRKSF--PCEQGLDNGYSISKFNHKGQGEYIFHAAENDDAQFTKM 431
QY 439 LPTVGDASACLCPADNTRCATNTGYMLS-----GGLCKPEVAESTDIHYIGFE 489
DQ 432 PTNIRKKPKOVIAFASASQASC--FSDGYPLSWVWKKCDKSPNCTREHTEGVNM----- 484
QY 490 TDLQLEMKYLLQKTPPTFVIAIFTSNDMPLENSWFDPSWPKPMILLTK--SNKYFSSSVH 548
DQ 485 -----PKANKV-----PQWVSSS-----TLNSEAATKGLV 513
QY 549 MGLGSLQICLTNS--TIEFWLVAVVNPFGCSSESFMVNVNSPDMWPKIKIDPL 605
DQ 514 -----GCAYNSLSTISCTILLSPGPF-----PFLUNISFVATIAV-L 553
QY 606 QYNWTLTANKKTKFTETVHYLSPKSPGPNFISY-----YE-PLFIDPSRNLG 659
DQ 554 FIVVITLILCKYKKQKPP--YEQIQMVGVVTCSSINERYVYVIRPKYDLKWEFFPENLE 611
QY 660 YMKINNTQVGRYSMH 674
DQ 612 FKKVLGSAFGKVM 626

RESULT 14
US-08-476-515A 84
: Sequence 84; Application US/08476515A
: Patent No. 6239270
: GENERAL INFORMATION:
: APPLICANT: Akorstrom, Goran
: APPLICANT: Jublin, Claes
: APPLICANT: Bask, Lars
: APPLICANT: Crumley, Greqg R.
: APPLICANT: Morse, Clarence C.
: APPLICANT: Murray, Edward M.
: APPLICANT: Hjalms, Goran
: TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
: NUMBER OF INVENTION: Thereof and DNA Encoding Same
: NUMBER OF SEQUENCES: 84
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Martin Savitzky
: STREET: Rhone Poulenc Rorer Inc., 500 Arcola Rd.
: STREET: 3C43,
: CITY: Collegeville
: STATE: PA
: COUNTRY: USA
: ZIP: 19426-0107
: COMPUTER READABLE FORM:

```

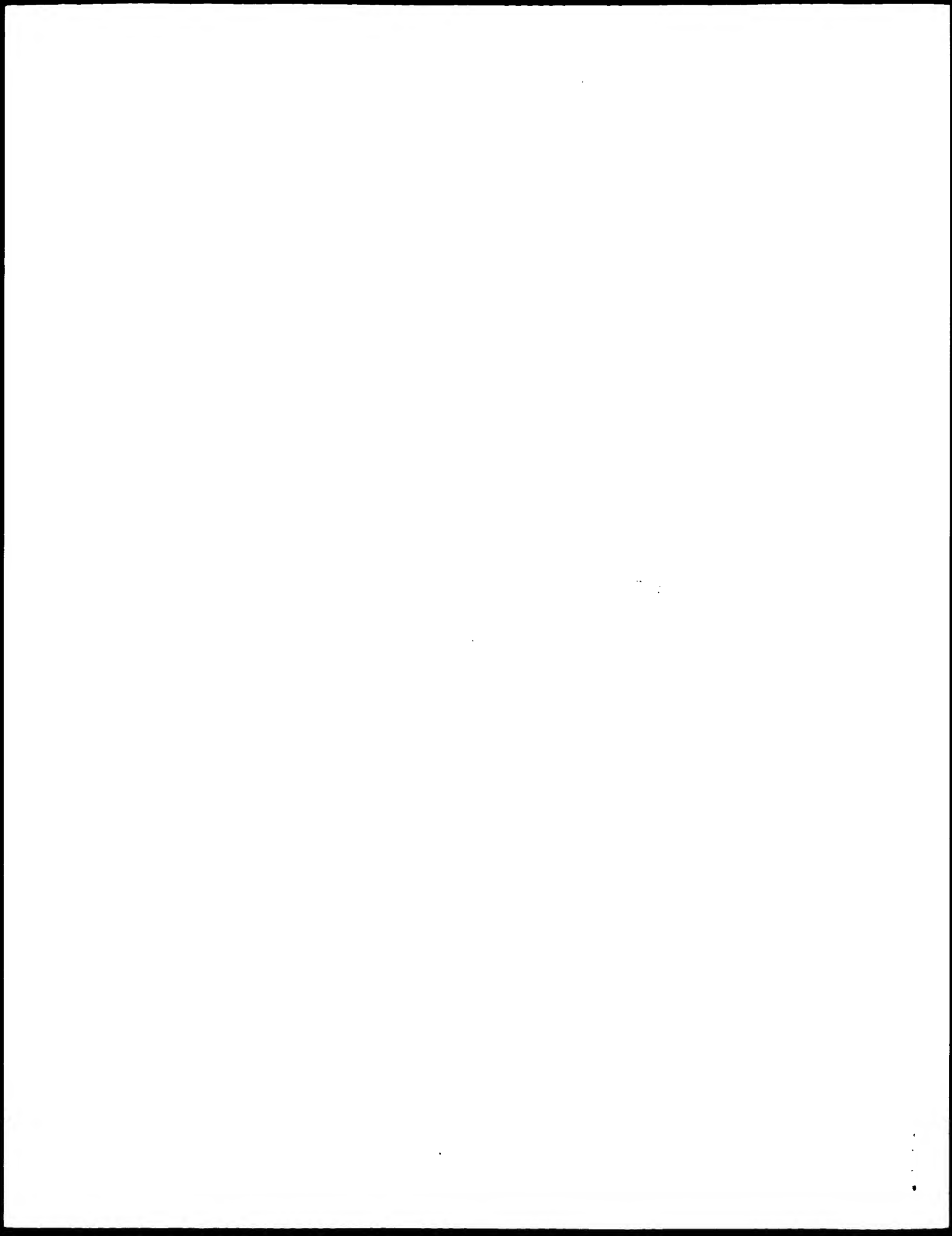
```

MEDIUM TYPE: Floppy disk
COMPUTER: Compat PC
OPERATING SYSTEM: Windows 95
SOFTWARE: Word 7.0 (Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,515A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,836
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/36,747/00483
FILING DATE: 24-MAY-1994
FELAS APPLICATION DATA:
APPLICATION NUMBER: SE 9301764-8
FILING DATE: 24-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Savitzky, Martin
REGISTRATION NUMBER: 29,694
REFERENCE/AGENT NUMBER: A1355D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610 454-3816
TELEFAX: 610-454-3808
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 4654 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-476-515A-84

Query Match          2.5%: Score 101.5; DB 4; Length 4654;
Best Local Similarity 20.3%; Pred. No. 40;
Matches 107; Conservative 50; Mismatches 183; Idels 187; Gaps 31;

QY 268 PPFCKPKNWCHGCKPKPECNQISMDIAMFNILPILFIWA-----YNSDPFES 420
DQ 1708 KWFVCSNSQAVV--PTPECMCHSEDEFYCEGIC--FENLWEGHDFWLYSGD 1261
QY 321 DPFCKMKPLPMNYFI--NTSTIMILWMSNFGPPYFALENSMPKLELKAQIVHLEF 478
DQ 1262 EHNACVYKICISYFCHDHCNCIHRANLCGR 5100
QY 479 LKPPCHVQPIISPPAPTSTYW-----LTPCTSTY-----GSEHLLG 417
DQ 1301 DEKCGTQFT--PFFSWQGLSHINI-VNLAVVLSLLELLENGLHESLGN--NS--SD 1355
QY 418 SPSEETHSCF-----CP-----NIOVVCYAFICIV-CHASA-CLICADPNRTG 460
DQ 1356 FNGGCTHECVQEEFKAKCLCPGLFLLANISKTCIDEDIDGSSSQHCYMR--GSEFC 1413
QY 461 GTGNTGYML--SQG--LCKPEVAEST-----CP-----DHYIGFE 489
DQ 1414 -SDTGYMLFSGMPTCKVTASPSILLIYASQNKIIADSVISQVNTYSLVNCYSYIVAV 1472
QY 490 TP-----LQDNEMFYLLQYTPPPFVIAFTISNEMFLNSWTF--WPKPMILT 540
DQ 1473 FDSISGRIFWSDATOGKITSAPFNGTDRFVDFSSILLITETIAHWVGRNLYWIDYALE 1542
QY 537 LKSNKYKSSLVHMLGLSLQICLTKNSTLEPVLAVVNPFGCSSESFMVNVNSPDMW 596
DQ 1543 IEVSK-----IDSHPTVLISKNLINPECLA--LLEPMNEHLEF 692
QY 597 -----ERIKDLPLQCYNWTITLGN--WKTFITVHYLSPKSPGPNESYIYEP 648
DQ 1574 GRHPRIERASMAQSMK-----IVIVQDKTFW-----PGLLIDY 1607
QY 649 LEFIDPSNLYAMK--INNIOVEGYSMHFDPEAL--POLLGLVYPT 692
DQ 1608 -----PNRLLYEMDSYLDYDFCYNGHIRKOVIASDLIIRPYALI 1649

```


```

Db 181 AASVFIDRSTLPLLIHIIQIASTAIKVTETPTGSLGCSNYENLESVSSVSVCSFENKIQI 240
QY 241 QCLQVLIPIYLRVQAAI SV AASVLI ELIKRIETETGCKKPPKPPNPNPNSMIDAMF 300
Db 241 QCLQVLIPIYLRVQAAI SV AASVLI ELIKRIETETGCKKPPKPPNPNPNSMIDAMF 300
QY 301 ENLLRITETWKAYNSDFEESDEKFLMKFLPMNYFLNTSTIMHLWTMDSNFORVEGLEN 360
Db 301 ENLLRITETWKAYNSDFEESDEKFLMKFLPMNYFLNTSTIMHLWTMDSNFORVEGLEN 360
QY 361 SMKQLEKAKIVHKLFSLSKKCHKQPLISIPPOPTSTYWLTPIQSELYVNEHILLASF 420
Db 361 SMKQLEKAKIVHKLFSLSKKCHKQPLISIPPOPTSTYWLTPIQSELYVNEHILLASF 420
QY 421 EPTHSTCTNDQVVCYAFAPICVQASACLTICAPINRRCGCHWGYMISGICKPPVAE 480
Db 421 EPTHSTCTNDQVVCYAFAPICVQASACLTICAPINRRCGCHWGYMISGICKPPVAE 480
QY 481 STDHYIGFTDQLEMKYLLQKTDRIEVRHAIFISNDRLNSWFDPSWKRKMLLTILKSN 540
Db 481 STDHYIGFTDQLEMKYLLQKTDRIEVRHAIFISNDRLNSWFDPSWKRKMLLTILKSN 540
QY 541 KYKSSLVHMLIGLSIQCLIKNSLLEPVLAVYVNPFGSSHSKSPMPVNSFPDWERIK 600
Db 541 KYKSSLVHMLIGLSIQCLIKNSLLEPVLAVYVNPFGSSHSKSPMPVNSFPDWERIK 600
QY 601 LKQYVWVLIQSKWAFELFVAVI KSKKASGNGNSESIVYERQEFIDSRNIGY 660
Db 601 LKQYVWVLIQSKWAFELFVAVI KSKKASGNGNSESIVYERQEFIDSRNIGY 660
QY 661 MKTNNTOVCYSMIHDPENAIPLIILQIDYPTQSGDSALLQILFTDPVKNLSPPQRP 720
Db 661 MKTNNTOVCYSMIHDPENAIPLIILQIDYPTQSGDSALLQILFTDPVKNLSPPQRP 720
QY 721 LDLFSCLLRHRLKLTSEVVRVQSALQAFNAKLPNTMDYDTIKLCS 766
Db 721 LDLFSCLLRHRLKLTSEVVRVQSALQAFNAKLPNTMDYDTIKLCS 766

RESULT 2
US-09-864-761-43251
: Sequence 43251, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 2426336
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30

```

```

: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 43251
: LENGTH: 378
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL022143.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
: OTHER INFORMATION: EST_HUMAN HIT: AW955819.1, EVALU0.00e+00
: OTHER INFORMATION: SWISSPROT HIT: Q28661, EVALU0.00e+00
US-09-864-761-43251

Query Match 38.0%, Score 1550; DB 10; Length 478;
Best Local Similarity 74.7%, Fred. No. 9.6e+27;
Matches 278; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 395 RSTYWRTRIOSYVYCNENGLIGSHSSEKTHSCNPOVVCIAFLPCTVGVASACLTICAP 454
Db 7 RSTYWRTRIOSYVYCNENGLIGSHSSEKTHSCNPOVVCIAFLPCTVGVASACLTICAP 454
QY 455 DNRTGCTGTGMYLSQIKKPPVAESTDHYIGFTDQLEMKYLLQKTDRIEVRHAIF 514
Db 67 DNRTGCTGTGMYLSQIKKPPVAESTDHYIGFTDQLEMKYLLQKTDRIEVRHAIF 514
QY 515 ISNDRLNSWFDPSWKRKMLLTILKSNKYKSSLVHMLIGLSIQCLIKNSLLEPVLAVYV 574
Db 127 ISNDRLNSWFDPSWKRKMLLTILKSNKYKSSLVHMLIGLSIQCLIKNSLLEPVLAVYV 574
QY 575 PFGSSHSKSPMPVNSFPDWERIKLIDPQVYNWVLIQSKKASGNGNSESIVYERQEFIDSRK 634
Db 187 PFGSSHSKSPMPVNSFPDWERIKLIDPQVYNWVLIQSKKASGNGNSESIVYERQEFIDSRK 634
QY 636 SNTPNFTSYVEPLEFTDPSNIGYMPINNVYVYSMIHDPENAIPLIILQIDYPTQSGDSALLQILFTDPVKNLSPPQRP 694
Db 247 SNTPNFTSYVEPLEFTDPSNIGYMPINNVYVYSMIHDPENAIPLIILQIDYPTQSGDSALLQILFTDPVKNLSPPQRP 694
QY 695 SCESALLQILFTDPVKNLSPPQRPGLLSNLLRHRLKLTSEVVRVQSALQAFNAKLP 754
Db 307 SCESALLQILFTDPVKNLSPPQRPGLLSNLLRHRLKLTSEVVRVQSALQAFNAKLP 754
QY 755 NTMDYDTIKLCS 766
Db 367 NTMDYDTIKLCS 766

RESULT 3
US-09-864-761-43252
: Sequence 43252, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 2426336
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30

```


Db 855 LSIIRFMMKLSLQDQKLIKLNLDG--KHKLK 886

RESULT 9

US-09-991-496-22

; Sequence 22, Application US/09991496

; Patent No. US20020169285A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, David C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-11-20

; NUMBER OF SEQ ID NOS: 137

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 23

; LENGTH: 320

; TYPE: PRT

; ORGANISM: Leishmania major

US-09-991-496-22

Query Match

Best Local Similarity 32.8%; Score 114.5; DB 9; Length 320;

Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;

QY 383 CHKQPLISLPRTSTYWLTRIQSELYNENGLLSFSEETHSTCPNDQVVCIAFLPCT 442

Db 176 CRISDACSVPNCKKCKFTGTSPL-----CAEDIGYSLALISCSPTTQ-----PCE 223

QY 443 VGDASACLTCAPNRRCCTGTCNTGYMISOGICK 475

Db 224 V---EHCNTCVNGDSTRCAVCNTGYVYVSDGCK 253

RESULT 10

US-09-991-496-55

; Sequence 55, Application US/09991496

; Patent No. US20020169285A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, David C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-11-20

; NUMBER OF SEQ ID NOS: 137

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 55

; LENGTH: 320

; TYPE: PRT

; ORGANISM: Leishmania major

US-09-991-496-55

Query Match

Best Local Similarity 32.8%; Score 114.5; DB 9; Length 320;

Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;

QY 383 CHKQPLISLPRTSTYWLTRIQSELYNENGLLSFSEETHSTCPNDQVVCIAFLPCT 442

Db 176 CRISDACSVPNCKKCKFTGTSPL-----CAEDIGYSLALISCSPTTQ-----PCE 223

QY 443 VGDASACLTCAPNRRCCTGTCNTGYMISOGICK 475

Db 224 V---EHCNTCVNGDSTRCAVCNTGYVYVSDGCK 253

RESULT 11

US-09-874-923-22

; Sequence 22, Application US/09874923

; Patent No. US20020081320A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, David C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-06-04

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 22

; LENGTH: 320

; TYPE: PRT

; ORGANISM: Leishmania major

US-09-874-923-22

Query Match

Best Local Similarity 32.3%; Score 114.5; DB 10; Length 320;

Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;

QY 383 CHKQPLISLPRTSTYWLTRIQSELYNENGLLSFSEETHSTCPNDQVVCIAFLPCT 442

Db 176 CRISDACSVPNCKKCKFTGTSPL-----CAEDIGYSLALISCSPTTQ-----PCE 223

QY 443 VGDASACLTCAPNRRCCTGTCNTGYMISOGICK 475

Db 224 V---EHCNTCVNGDSTRCAVCNTGYVYVSDGCK 253

RESULT 12

US-09-874-923-55

; Sequence 55, Application US/09874923

; Patent No. US20020081320A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, David C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-06-04

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 55

; LENGTH: 320


```

: CURRENT FILING DATE: 2001-06-27
: PRIOR APPLICATION NUMBER: 09/245,041
: PRIOR FILING DATE: 1999-02-05
: PRIOR APPLICATION NUMBER: 60/093,630
: PRIOR FILING DATE: 1998-07-21
: PRIOR APPLICATION NUMBER: 60/104,978
: PRIOR FILING DATE: 1998-10-20
: NUMBER OF SEQ ID NOS: 129
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 17
: LENGTH: 1350
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-893-238-17

Query Match      2.5%; Score 100.5; DB 10; length 1350;
Best Local Similarity 20.68; Pred. No. 18;
Matches 70; Conservative 34; Mismatches 76; Indels 159; Gaps 20;

QY 253 ERFVQAALSYIA'CNSEGE-FTG--KENDG-WF--HGGPKPEEC-----NCP--- 292
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 690 ECFSKRTLDHRCDDHTDCYSCYCTANTNDCHWCNDHCVPRNHSCSGOISIFRYENCCKDN 749
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 293 -----SMDIQAMEENL-----LRITETWKAYNS 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 750 PMTYCNKKTSCRS'ALDQNCWEPPNQE'IALPENIT'G'IMHILV'NSC'LKITAKENY-- 807
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 DFESDEPKLPMKRLPMNYFLNTSTIMHLWTMDSNFORRYHQLFNSMKQLPIKAOKIVHK 375
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 808 -----DNAKLFERN-----HALLASLITQ-----KKVEFVLKQL-----RIMQS 842
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 376 LFSLSKRCHKQLISLPQRTSTYW--LTRIQSELYCNE-----G 414
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 843 SOSMSK-----LTIITWVGLRKINVSYWGWEDMSPTNSILLOWMPSEPSDAG 889
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 415 LLGSESE-----ETHSCTCPNDQVVC-----TAFLECTVGDASACL 450
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 890 FCGILSEPSTRG'KAATCINPLNKSVCDEPPANHSKAKQCP'CALRTACGDSCTSG--SSECM 948
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 451 TCAPIDNRTHC-----GTCNIGYMI'LOGICKPHE 477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 949 W'S--NMKQCVDSNAYVASFPFGQCMWYIWS--TCPPE 983
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: May 19, 2003, 10:35:44
Job time : 34 secs

sequence version 1.1.4 (1.4578
date: 09/1993 2003 comparison: 1td

carcin using sw model

2003.10.28.53 : Gen to file 47 Seconds
(without alignments)
1995.787 Million cell updates/sec

seq 129 45

AKA:AELESLMALWEW : LAINAKLPNTMDYITKLCS 766

seq 129 45
seq 129 45

seqs: 9614422 residues

displaying chosen parameters 283224

2003.10.28.53

am March 0K

am March 100K

for first 45 Summaries

11

11

11

11

number of results predicted by chance to have a
in or equal to the score of the result being printed,
analysis of the total score distribution.

SUMMARIES

seq	length	IP	Description
1	2	100052	hypothetical prote
2	2	852835	hypothetical prote
3	2	144178	large tegument pro
4	1	A27840	complement C7 prec
5	2	144991	large tegument pro
6	2	S46774	GTPase-activating
7	2	140644	botulinum neurotox
8	2	S47928	signal recognition
9	2	148755	hypothetical prote
10	2	A44257	interleukin-6 sign
11	2	128481	ORF MSV240 leucine
12	2	T32276	hypothetical prote
13	2	141934	large tegument pro
14	2	S45734	probable membrane
15	2	A47182	perforin 1 precurs
16	2	T13029	1pr homolog - trul
17	2	T15689	hypothetical prote
18	2	T48842	probable RAS GTPas
19	1	14083	integrin, band 3 p
20	2	E32734	conserved hypothet
21	1	101759	genome polypeptin
22	1	T4558	hypothetical prote
23	2	D62891	exodeoxyribonuclea
24	2	A45816	cytolysin precurs
25	2	143274	dynamin heavy chain
26	1	A42399	isoleucine-tRNA li
27	1	147459	keratin Ha3-11 ty
28	2	296270	hypothetical prote
29	2	S47202	probable membrane

ALIGNMENTS

RESULT 1

T09052

hypothetical protein DBCCR1 - human

C:Species: Homo sapiens (man)

C>Date: 11-Jun-1999 #sequence_revision: 11-Jun-1999 #tax_change: 21-Jul-2000

C:Accession: T09052

R:Rabuchi, J.; Luscombe, M.; Elder, P.A.; Knowles, M.A.

Genomics 48, 277-288, 1998

A:Title: Structure and methylation-based silencing of a gene (DBCCR1) within a candid

A:Reference number: Z16537; NCBI:98207242; PMID:9545642

A:Accession: T09052

A>Status: preliminary; translated from CH/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-761 <HAR>

A:Cross-references: EMBL:AF027734; NID:q3041876; PID:AKC9691.1; PID:q3041877

A:Experimental source: tissue type: brain; developmental stage: fetal and infant

C:Genetics:

A:Gene: DBCCR1

A:Map position: 9q32-q33

A:Note: H3089A

C:Superfamily: human hypothetical protein DBCCR1

Query Match 53.0% Score 2161; DB 2; Length 761;
Best local Similarity 52.3%; Pred. No. 1,16-148;
Matches 406; Conservative 114; Mismatches 209; Indels 28; Gaps 10;

QY	1	MWRSRAGAELESLMALWEWIALS-LHWLAVAAVSRQHATSI	FWLLISLQWGFHPSQE	59
DB	1	MNRRE---VELLYFLWGRISVQPSH---	DEPAGTIDHVSKEFWLLISLQWGFHPSKS	53
QY	60	YIIPVMSWGQSTPKIYFPGRWKVNIAVERNNFAGSPHIAPEF	FNRIKGRPT	119
DB	54	YLSEVHEHFGFTYKRYFAEAKVRNTATERRDLVHPVPMHET	QRSLDRPT	113
QY	120	LOQITENLKFYTHRELLSAILDHFPSLI	FVPRFPIKFAEISISTNSSSVILETRQ	179
DB	114	TQFPIPLKKGTHLHLSAILGGSLALIMYKSKM	-----LAKSONATQSVLEALHQ	166
QY	180	LAASYFIDRSTLRRLHFIQIASTAKVITFTTGPISNYNI	IVSVSVLSQSPENKIQ	249
DB	167	LASSYFVDEGIMPRFHTQISTGALKVETETGTGLYSNSYNI	IVSVSVLSQSPENKIQ	226
QY	240	LOQVLAALYQERVAALSYIAGNSGDF	LEKNDWCHQOKFPGNTPSMQIAM	299
DB	227	LOQLQLFFYLQERVQSALSYMNGEYDQNSQGRQWAFEP	QWNPFTIDQIM	286
QY	300	EENLLRITETKAYNSDFESDEFFKFMKRLPMVFNLTSTIM	LWTMCSNFORVAYOLE	359
DB	287	EYLANMAKSWALAYKLENSDFKSFPMKRLSNHPLIGUS	INLHWGNWLOLONKYKLO	446
QY	360	NSMKQLFLKAKIVHKIFSLSKPCHKQPLISLFRPMTSTY	WLTEIQSTLYNENGLASF	419

Db 347 SATPAQRQKIQRTARKLPGISVRCRHNPNHQIPRRTIQOULARVOSILLYCNEGFWGTF 406
 QY 420 SEETHSLTLENDQVATAFLEPTVIGASALITAPGNTRPGTNTGMYMLSGLEKKEPVA 479
 Db 407 LESQSCVWIGSTILQRPPIVIGNNNSCTMCSLANISLQSCNKGKLYRGREPNV 466
 QY 480 ES--TDHYIGFDTL--QDLEMKYLLQKTRDRIEVHAIFISNDMRINSWFPDPSWRKRMILG 535
 Db 467 DSRSEQFISFETDLPQDLEKLYLQKMDSRLLYVHTTIFISNEILDTEFDPWRKRMSL 526
 QY 536 TLKSNKYSLSVIMILGSLQILCKNSTLEPVLAVYVNPFGSSISESWFPVNPENSEPD 595
 Db 527 TLKSNKNMDFIMVTCMSMRICOMNSSLDPMFFVYVNIPTSGSHSEGNMPPFORFGYPR 586
 QY 596 WERTKLDLPQCYNTWLTGNCWKTFEFVHYIYLSRSKNSGPNCESTYEPELEFIDPS 655
 Db 587 WEKIRLQ--NSQCYNTWLTGNCWKTFEFVHYIYLSRSKNSGPNCESTYEPELEFIDPS 644
 QY 656 RNIGYMKINNIQVFGYSMHFDPPEAIDQLLDYPTTQGSQSSALQLLEIRQ 711
 Db 645 KPGFYKISVQVQVGYSLDFENALLPSAVQVQVNSYTGQGFYSSSSVWMLLLLTIPCRIN 704
 QY 712 KLSF--EFGPRIDTFCSTLPRKLLKSTSEVVRQIQAQENAKLPNTMDYDTTKLC 765
 Db 705 KIAPPVAKGPKQDLDLPSMLKRRKILINSEIRVNHALDLYNTHIKOSDQMTAKLC 761

RESULT 2

S52835
 hypothetical protein, YMK075w, yeast (Saccharomyces cerevisiae)
 N: Alternate names: hypothetical protein YMK075w, YMK075w
 C: Species: Saccharomyces cerevisiae
 C: Date: 10-May-1995 #sequence_revision: 61-Sep-1995 #text_change: 1-Apr-2002
 C: Accession: S52835
 R: Pearson, D.; Bowman, S.
 submitted to the EMBL Data Library, April 1995
 A: Reference number: S52814
 A: Accession: S52835
 A: Molecule type: DNA
 A: Status: preliminary, translated from cDNA/EMBL/DBJ
 A: Cross-references: EMBL:248452; NID:g763008; PIDN:CAAR8800.1; PID:g763022; GSPDB:GN0001
 A: Experimental source: strain AB972
 C: Genetics:
 A: Gene: MFS:YMK075w
 A: Cross-references: SGD:S0004680
 A: Map position: 13R

Query Match 3.38; Score 135.5; DB 2; Length 684;
 Best local similarity 18.68; Pred. No. 0.1;
 Matches 128; Conservative 86; Mismatches 221; Indels 255; Gaps 36.

QY 76 KIYPERGRKVVNLAVERPNTFGSLPLAP-----EFPNIRLLGSPPTLQQTENLIK 129
 Db 122 KIKRE-SIMNYRKNLCQSN-----SMHTLVPSKRPQTQVKNPQDINRNDIKTFITENMT 177
 QY 130 KYCTHPLISATIGGCKSLITFVDPKKIKSKRGSDSTTSSSVTLETILHQLAASYFIDRD 189
 Db 178 ESN-----IPSTICWQDIT-----NATROPESDPEDNKKLSNIRTKIILSNATY----D 225
 QY 190 STLRKHHLHIQIASTAIKVITRIKPIGPGSNYINIDSVSVLQSPENKIQIQAQVLLDPO 249
 Db 226 SKSLFQCSNKSSTS-----NASHKIFPKDKNNS-----TID 256
 QY 250 YLQKRFQAALSYIACNSGEEELKENDQWHLGKFFECNC--PSMDLQAMEENLRIT 307
 Db 257 FNRHDFCS-----ACNQSGLPIC-----CUTCPKSHFICLDPPIDPNNIKPDMDHCN 304
 QY 308 ETWKAYNSDPESDEPKLPMKRLPMNYFIINTSTIMHLTWDSNFORRYEQLENSMKQLPL 367
 Db 305 EC-----KFKIFI-----NNSMATLKKIESNFIK-----QNNNVKIFA 337
 QY 368 KADQIVHKLISLISKKCHKQPLISLPQRTISIVYILTRIOSPIYCNINGILGFSSETHSCT 427

Db 338 KL-----LPNIDSHNPKO--FQIDPNYIKETPPAVK-----TQSRGOYSUF----- 375
 QY 428 QNDQVWVCTAFLEPCVIGASACILCAPGNRTRGCTNENIGYMLSGLEKKEPVAESTDHYIC 487
 Db 376 --NDKRIPLT-----LFTSY--QGSITK-----LDSY-N 402
 QY 488 PFTQILQILQEMKYI LQKTRPPIPVHAIFISNDMPINSWFPDPSWRKRMILITKSNPKYSLSV 547
 Db 403 PDTHIDNSGKELI-----CYKQNTRLGWSIPE--NSRLIMT----- 449
 QY 548 HMTLSLSLQIGTCTKNTSTLEPVLAVYVNPFGSSISESWFP--VNSNSPDPWEPTKLDLPLQ 606
 Db 440 -----CUCQCT-----PWHLDQVPRASPKN----- 459
 QY 607 CYNWTLLTGNKKK-----TFEETVH-----IYLPSPKSKSGPNESISYIEP 648
 Db 460 -----LASKKKCPHLSPTKYKKIHHQEDNSVNYFVWKQKQL-----INFRKMLIYEP 508
 QY 649 LEFIDPSRNLGYMKINNTQVFGYSMHFDPPEAIDQLLDYPTTQGSQSSALQLLEIRQ 708
 Db 509 LQ-----KIGYCNNGNQLIETTSITDQD---PQDCFKITQIDENS:KYDEF---D 553
 QY 709 RVNKLSPPGQPPLELFLSCILLRHLPLKLTSE 738
 Db 554 KIYKSKMWOKRKLPQFQESLIDKIVNSGQ 583

RESULT 3

T44178
 large tegument protein U31 [similarity] human herpesvirus 6 (strain 229)
 C: Species: human herpesvirus 6
 A: Variety: strain 229
 C: Date: 21-Jan-2003 #sequence_revision: 21-Jan-2003 #text_change: 05-May-2000
 C: Accession: T44178
 R: Dominguez, G.; Pambouk, T. F.; Stamey, F. P.; Dewhurst, S.; Inoue, N.; Bellett, P. E.
 J. Virol. 74, 8040-8054, 1999
 A: Title: Human herpesvirus 6 genome sequence, coding content and comparison with human
 A: Reference number: 222734; MUID:93412318; IMID:10482553
 A: Accession: T44178
 A: Status: preliminary, translated from cDNA/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-2077 <DOM>
 A: Cross-references: EMBL:AF157706; PIDN:AA049645.1
 A: Experimental source: strain 229; variant B
 C: Genetics:
 A: Note: U31
 C: Superfamily: varicella-zoster virus 240K tegument protein

Query Match 3.28; Score 129; DB 2; Length 2077;
 Best local similarity 21.08; Pred. No. 1.5; 248; Indels 292; Gaps 44;
 Matches 172; Conservative 109; Mismatches 248; Indels 292; Gaps 44;

QY 57 SQEYTDVFIQSPGFTSTPKYIYRFGSPKVNLA-----VEPRNFGSPPLAPEFFRNI 111
 Db 584 SEKLINTIHTLNKKT---KIVQD-----NYNIAAGFMVTEINCHAYI---VNLYFNI 633
 QY 112 KILGR-----KPLDQITENL-----IKKYGT-IF-----LLSAILGSESLTIVDKRL 156
 Db 634 EVLKHSGNLINTELDQVEKLYDNTQFLRFGTSHFNINLSNFTLSIRKMEVDYNSQKP 693
 QY 157 SKRPG-----SDSTNSSSVTLETILHQLAASYFIDRDSTIRPHHIGIASTAIKV 207
 Db 694 SDPASEIIAIESILADPSKNTIVNIEMI-----KSQLEELGKMEISTENKQ 742
 QY 208 TTRTGP-----LQCSNYDLDSVSS-----VLVQSPF-----SKIQI 240
 Db 743 TAFITKQILGQQLPIIVDFLHLSAYNLPNTITVKNLHIFILEKRPDIAAIIHDKIQ 801
 QY 241 LSLVLIPIYIQEPFV--LQALSYIA--NSFDEL--KENLW--HGFKE--NFK--MFCAM 299
 Db 802 SLIDICIDMLNDIIVPEQTSTVL-----ELV-----DLFEN-----STERTAL 841
 QY 300 EENLLRITETWK--AYNSLFESEDFKLFMKRLPMNVFIINTSTIMHLTWDSNFORRYEQ 357

A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A and
 A:Reference number: 225732, MUID:9912319, PMID:10482554

A:Accession: T439a1

A:Status: preliminary, translated from GB/EMBL/TrEMBL

A:Molecule type: DNA

A:Residues: 1 2077 (15%)

A:Cross-references: EMBL:AH62150, NID:94495477, PIDN:AAA78252.1, PID:91960019

A:Experimental source: strain HST, pop. variant B

C:Genetics:

A:Note: U31

C:Superfamily: varicella-zoster virus 240K tegument protein

Query Match 3.0%; Score 124; DB 2; Length 2077.

Best Local Similarity 20.8%; Pred. No. 35;

Matches 171; Conservative 110; Mismatches 248; Indels 292; Gaps 44;

```

QY 57 SOEYDFVDSRQGFSTRKYIYFEGPWKNVLA-----VEPRNELSGPLPLAPEFFPNT 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 SEKITNTHTLRNKPT---KIVOD-----NYNLAAGFMPVTEHLCIFAYL---VNLXFN 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 112 PLLGP-----PPTLOOTEM:-----IKRYGT-HF-----LLSATLGGEESLTIFVDKEKL 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 EVLKHSGLNINTELLQVEKLYUNTQFLPFGSHENINNLNFTLSIRKMEVDFVNSQKP 693
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 157 SKRAG-----SDNTSNSSSVTLETHOLAASYFIDRSTLRRLHHIQIASTAIKV 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 SDRASEILAAIESILADPSKNKTIIVNEM:-----KSLEELGKMEISTENKQ 742
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 TETHTGP-----LRSNYNILNLSVSS-----VLQSPF-----NKIQQL 740
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 TARIKTKQILGQDELTPIDYDLHLHSAYNLPNTTIVKNLHLHLILKRPDIAALLHDKIQ 801
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 QGLQVLPTVLEFFV--QAALSYIA"NSRSEFTIKENLW"HP"PKFPEON"PSMLQAM 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 SILDICIDMDLNTVPEOTFSTVL-----FLV-----DLFPN---STEKIAL 841
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 300 PENILRITETPW--AYNSDFESEDREKLFMKRILPMNYPLNTSTIMHLWTMSDNFORRYEO 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 842 FESVILTPOLAKKCANLKTLEPFD-----LAQFITTS-----EQ 877
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 LENSMLKOLFKAQKIVHKFLSLSKRCHKQPLISLPQRTSTVYLTPIQSFLYCENGLLG 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 LQNMKKOHF---CKKIPTLMHDIKPLYSQKIITAEKN---WIGR----- 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 418 SFSEETHSCTPNDVVCTAFLEPTWEDASACLTAPDNKTHGTCNTGYMLSGCLAKPE 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 ---AKTAVITSPPE---LTAFL-----ATAPTKHALQ---TCAPE 947
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 478 VAPSTHVIIGFETNQLQEMKYIILQKTRPPIEVHAIIFSN--DMRINSWFDPSWKKRMIL 535
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 948 LKALOPH-----MEFOMKQTAENDKKHILTPNTLEKPLND-----III 987
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 TLKSNKYK-----SLVHMILGLSLQICLTKNSTLEPV 568
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 LKNGQPSSEFVHNLNLETLKQIQUNDIITPHHALIPVTKIETWISKFISIDLEKI 1047
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 569 L AVYVNPFGGSHSESWEPVNSFPDWERIKLDLPGLCYNNKWTLTGNKW-KTFFETVH 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 LIKTPLNPEQMSBEQKYTPL-----LSFLSKFKKTKTECT-- 1082
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 627 IYLRIRKSNCPNGNSIYVEPIEFIDPSRNLGYMKINNIOVFGYSMHFPDPAIRDLILQ 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1083 ---EDVKTEIDQMKSTIF--LTKIATSTN-KYTRISH-SVYGQELNLYEERITELKKE 1134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 687 LDYPTTG"SSJ"ALLQ---LLEIKIKI-PVVKLSFEP"PPHILPS"LLPHPL----- 742
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1135 TNKIKEPLSKREYAVAKKKIILSSQDAKYNK-----IYLVNTHHTTKETKNTQPRE 1184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 733 ----KLSTSEV----VRIQSALQAFNAKLPNTMDYD--TKL 764
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1185 TASAKALTVEVNNKENQLOELLNHNFAHLKAKDQNIHTKL 1225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

S36773

GTase-activating protein BMD2 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: cyclin CLA2; protein YKL020; protein YKL424

C:Species: Saccharomyces cerevisiae

C:Date: 03-May-1994 #sequence_revision 03 May-1994 #text_change 29-Oct-1999

C:Accession: S36773; S37917; S37918; S41294; S39117; S42005; S34706

R:Park, H.O.; Chant, J.; Herskowitz, J.

Nature 365, 269-274, 1993

A:Title: BMD2 encodes a GTPase-activating protein for Rho1/Rsr1 necessary for proper
 A:Reference number: S36773; MUID:93982538, PMID:8371782

A:Accession: S36773

A:Molecule type: DNA

A:Residues: 1-1104 (CLAP)

A:Cross-references: EMBL:L19162; NID:9402733, PIDN:AAA34461.1, PID:9402734

R:Pohl, T.M.; Pohl, F.M.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S37897

A:Accession: S37917

A:Molecule type: DNA

A:Residues: 1-681 (POL)

A:Cross-references: EMBL:228092; MIPS:YKL0920

A:Experimental source: strain S288C

R:James, C.M.; Gent, M.E.; Indge, K.J.; Oliver, S.G.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S37918

A:Accession: S37918

A:Molecule type: DNA

A:Residues: 237-1104 (TAM)

A:Cross-references: EMBL:228092; MIPS:YKL0920

A:Experimental source: strain S288C

R:Yurokova, P.; Nasmyth, K.

EMBO J. 12, 5277-5286, 1993

A:Title: Yeast G(1) cyclins CLN1 and CLN2 and a GAP-like protein have a role in bud 1

A:Reference number: S41683; MUID:94085403; PMID:8262070

A:Accession: S41684

A:Molecule type: DNA

A:Residues: 1-436, Y438-1104 (CVK)

A:Cross-references: EMBL:X74130, NID:9395287, PIDN:AAA52228.1, PID:9395289

R:James, C.M.

submitted to the EMBL Data Library, October 1993

A:Reference number: S39116

A:Accession: S39117

A:Molecule type: DNA

A:Residues: 237-1104 (JAZ)

A:Cross-references: EMBL:X75561; NID:9414687; PIDN:CAA53241.1; PID:9414689

R:James, C.M.; Gent, M.E.; Oliver, S.G.

Yeast 10, 257-264, 1994

A:Title: Sequence analysis of a 3.5 Kb EcoRI fragment from the left arm of Saccharomy

roten.

A:Reference number: S42004; MUID:94262329; PMID:8203166

A:Accession: S42005

A:Molecule type: DNA

A:Residues: 237-1050 (JAZ)

A:Cross-references: EMBL:X75561

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SUP-BUL2; CLA3; ERK25

A:Cross-references: SGD:S0001575; MIPS:YK10920

A:Map position: 111

C:Superfamily: ras-specific GAP catalytic domain homology

t, 520-741/domain, ras specific GAP catalytic domain homology ~GAP>

Query Match 2.9%; Score 118.5; DB 2; Length 1104;

Best Local Similarity 19.3%; Pred. No. 35;

Matches 134; Conservative 96; Mismatches 227; Indels 237; Gaps 31;

QY 109 RNRIILGRPTTQQI-----TENIKKYGTHFI-LSATIGGRSLTFV 151

Db 416 KRLKLGKIKITQEIINDTRYNKETRLPMDVDNKNFQIGTICIKISSNLFLPSTNFV 475

QY 152 DKKKISKRA-----EGSDSTTSSSVTLETHOLAASYFIDRSTLRRLHH 198

A>Note: 269.Scr was also found

R:Polh, I.M.

Submitted to the Protein Sequence Database, May 1996

A:Reference number: S64899

A:Accession: S65272

A:Molecule type: DNA

A:Residues: 1-599 <POW>

A:Cross-references: EMBL:Z73594, NID:01320500, PIR:Q246950, PDB:1J70501, MIPS:YPL243W

A:Experimental source: strain S28C (AB972)

C:Genetics:

A:Gene: SGD:SRP68

A:Cross-references: SGD:S0000164, MIPS:YPL243W

A:Map position: 161

Query Match 2.9% Score 117.5; DB 2; Length 599;

Best Local Similarity 19.9% Pred. No. 1.7;

Matches 72; Conservative 54; Mismatches 120; Indels 115; Gaps 15;

QY 87 NLI AVERNNFGSPPLADEFERNIRILGRPTLOOITENLIKYYG----- 132

DB 168 NNLAL-----VTALEHLKNSLLA-----EEVDNIVNKYQYSLKQYAGNLITTP 213

QY 133 --THFLSATLGE---ESLTFVVK---PKLSKRAEGSDSTNSS----- 170

DB 214 EINNFIKVERQSDNDDRLVRLLDNGFNMKKITTSTDDQKVTININWRSFNAKIIDAE 273

QY 171 -SVTLE---TLHQLAASYFDIPOSTLPLRL---HHQTASTATKVTETPTGPGGSNYDNL 223

DB 274 VAQFLEGLSIHTQITQYTRLSKLEKALDRIEFFIA-----NIDQ 316

QY 224 LSVSSVLVSGSPKNIQGLQVLVIVYQERFVQAALSVIACNSGPFCKENKCHQG 283

DB 317 DDIDEMVENSENN-----QILAYIKYNILLTSISRRDLFTHLW 357

QY 284 KPPECNCPSPMDIQAMEENLLRITETWKAYNSDPEE-----SDEF--KLPMKRLPMNYF 335

DB 358 NQWIKNTSLSPSKLIYKRMERIVKNIKYISDIMEIPGVYSDDRLQSGLDKLYQLP 417

QY 336 LNTSTIMILWTMSDFOR-----RYEOLN-----SMKQLFLKAKI--VHKLFSLSK 381

DB 418 LNTGTSVLVYQSGPYMEALVYVAYPPLENKLSIESLDEILLPALLSLNSVRSLOK 477

QY 382 R 382

DB 478 R 478

RESULT 9

T38755

hypothetical protein SPAC305.09c fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 04-Dec-1999 #sequence_revision 0; Dec-1999 #text_revision 0; Dec-1999

C:Accession: T38755

R:Oliver, K., Harris, B., Wood, V., Barrall, R.G., Rajandream, M.A.

submitted to the EMBL data library, September 1996

A:Reference number: Z21809

A:Accession: T38755

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2685 <OLI>

A:Cross-references: EMBL:Z99296; PIR:Q24694; 1; GSPDH:QND0066; SDBP:SPAC305.09c

A:Experimental source: strain 972h; cosmid C3H5

C:Genetics:

A:Gene: SPOR-SPAC305.09c

A:Map position: 1

Query Match 3.9% Score 117.5; DB 2; Length 2685;

Best Local Similarity 19.9% Pred. No. 15;

Matches 173; Conservative 124; Mismatches 263; Indels 309; Gaps 52;

QY 75 KYIYREGRMKVN NI AVERNNFG-----SPLPLAPEFRN 110

DB 1455 YDAYDFRSHYTHCSIALKSPRYLDANAFENVNSYNTIHILPLVFSFSSWNILFSNM 1514

QY 111 ---IRLGRRPPIQQIPNNIKKYSTHPIISALOGCHPSIIPV-----PKPKLSK 158

DB 1515 SYPLPSGNLFTPTDSSPN---KKFPHI---ATPKYALELIPISIMYNYKTNFNMWDR 1568

QY 159 RAPGSDSTINSSVILELHDLAASYFIDRINILPHIHH-----IQIASIAKVIETPI 212

DB 1549 TASATGIKAFVNFIIILHJSEKHEVKNKANLCPKHQFALSMMKVHLARIIPKILIDRA 1527

QY 213 -----GELGCSNYDNLDSV-----VQSPFNKIQIOLIV-----LLEP 249

DB 1628 ISASIDECAL-----DSDSEITANVLDDEEKFCPSFKNVDCIATWVDIDYHRAWLIPQ 1682

QY 250 YIQE-----RFOCAALSYIACN---SQCERI---CKENKCHGCGKPKFPHENC 291

DB 1683 QNEKSTYPIAPSPFIY---YHTRKHPNFKNKEIIPOTCRGDEFTI-----RCLM 1734

QY 292 PSMDIQAMFNELRITETWKAYNSDPEDEFKPKRKLPMNYFELNLSIMHFWIWSN 350

DB 1735 PSRENPKAVQYELIQ-----KKKRLLEPFM-----SS 1761

QY 351 FORRY-----EQLE---NSMKQLFLKAKIIVHKLFSLSKCHQPIISLPURTSTYWLTRI 404

DB 1762 RQRIQGLKSQLSNNDSREVRQRYEELIKRIVILSD---HYRILEYLIKDESSCSOASO 1818

QY 405 QSFVLCNENGLI---GSPSETHS---CICPDVQVCTAFPLCTVGDASACILCHD 455

DB 1819 -----CSENGQVLSYASLSVSVAHNNRFVHHNVQVKNNF1---RNVMYSYVIEV 1867

QY 456 NPIKQIPNINYMUSQ-----GCKPEVAHSTHYIGPFDIDOLF-----MKYLLOK 503

DB 1868 ERVR---GFAYMSQKALVRLDLKRTESANDOFFCONYTEDEDRENAPHLKLEED 1923

QY 504 TDRRIEVHAI-----FISNDMLNSWFEDSWPKRM 533

DB 1924 SKRRVWKTSDAIRHSGSPGNTSISNNPYDIQSIVVERFISPTQIGCSSNP---EKAV 1981

QY 534 LITLSKNYK-----SSLVHMIIGLSI-----QICLKNSLIEVIA---VYVN 574

DB 1987 TISTSLQIKLSVVIRPUNIVNITFPPITVKVNSDEFFISKPSDFVFNASSVLIN 2041

QY 575 PFGGSHSSEWFMVNFNSPFWEPKLDLDLQ---GVNMTLLGKSNKWIFFRIVHIVKRI 533

DB 2042 EYGCENITVW-----PPW-----VPFETEDVLIIPA---AFSRFLH-----RV 2077

QY 634 KSNPGNGNESIYEPELEFIDSPNLYMKINNIQVRSYSMIIFDEPAIKUL-----ILQLD 668

DB 2078 -----SFSVIYTKHNDRL---QPIVHSTRAPEDLDTHADILIPD 2115

QY 683 YNIGQSUSA---LQQLDLKQVKKLSHFGKRLKGLSGLRHKIK LSLSE 738

DB 2116 FPPVVFSTSSVYDIFITITLILYKEFAQF---GPNKPLEFIMIAHNSLISE 2170

QY 739 VVEFISLQA-----FNAKILNIMUDIT 762

DB 2171 VVR---ALQAVRHRLIDILKIQHQL---YDVI 2195

RESULT 10

A44257

interleukin-6 signal transducing molecule gp130 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Apr-1994 #sequence_revision 18 Nov 1994 #text_change 28-Jul-2000

C:Accession: A44257

R:Wang, Y.; Nesbitt, J.E.; Fuentes, N.L.; Fuller, G.M.

Genomics 14, 666-672, 1992

A:Title: Molecular cloning and characterization of the rat liver IL-6 signal transdu-

A:Reference number: A44257; M010:94052397; PMID:1427893

A:Accession: A44257

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-918 <WAN>

A:Experimental source: liver

QY 427 TCRNDQVCTAFLPCTGVDASAGLTICAPDNRTGCTGNTGVM--LSQGLCKPEVAEST--- 482
 DB 173 TLUDGLVITETVPCDQPGPTTVCVSQVCCAMNCCQCYGPMELKRNLTIAEIVPQAWYA 232
 QY 483 ----DHVIGFETDLOLEMKYLLQKIDRRLEVHAIFISNDMKLSWFPDPKRMMLTLK 538
 DB 233 PRLSHVEASITNVPMGMMEPEPHINDNVKAANTWDEN---GLIFSP--PK---VTVS 284
 QY 539 SNKYKSSIVHMLIGLSIOICITKNSITPPLVLAVYVNPFGGSHSSEWMPVNPSPDWER 598
 DB 285 TSGLVOLKRFKRESKALVALSNATIDV-----NNIMPIN-----RK 324
 QY 599 TKIMPLQVYNNWTLITLGNKKKIFFEIVHILYERKISNIPNESIYYEPELEFIIPSPNL 658
 DB 325 YKLSLLEFETLRGLSSKHKYKVLFFVMI-----ACVN-----DSMDAARL 366
 QY 659 GYM-----KINNIQVFGYSMHFDPHA 679
 DB 367 VELVQGIPOCKINLIQ-----FNPIIS 386

RESULT 13

T41933

large tegument protein - human herpesvirus 7 (strain J1)

C:Species: human herpesvirus 7

A:Variety: strain J1

C>Date: 03-Nov-1999 #sequence_revision 03-Nov-1999 #text_change 21-Jan-2000

A:Accession: T41933

R:Nicholas, J.

submitted to the EMBL Data Library, December 1995

A:Description: Determination and analysis of the complete nucleotide sequence of human H

A:Reference number: 223022

A:Accession: T41933

A:Status: preliminary, translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1:2059:NT*

A:Cross-references: EMBL:U43400; FJUN:AA054643.1

A:Experimental source: strain J1

C:Genetics:

A:Note: U31

C:Superfamily: varicella-zoster virus 240k tegument protein

Query Match 2.8% Score 113; DB 2; Length 2059.

Best Local Similarity 18.9% Pred: No. 22;

Matches 150; Conservation 38; Mismatches 248; Indels 216; Gaps 37.

QY 71 FSTRKYIYRFQKWKVNIYLAVERNNFLOSLPLAPDFRNIRILGCRPTLOQITFENL--- 127

DB 609 FAVLVKLYFDVYSITINGFVEN-----ELIKNI-----EQIYDNTQYL 647

QY 128 ---KKKGTHTPLSATGCTPFS--ETPVKPKIKSPAPGSGTINSSSVTIFTHQLAAS 193

DB 648 RPLGTRPMGNLTPTTISYKRWKHLDFPLSQKTLIDPAEETINLPFRKSVTPGKQKIATK 707

QY 184 YFDISTPLRLHHIQTASTATKVTPTGPGPGSNYP-----NLDSSVSVLVQ 232

DB 708 NMLREQ--LQJNAMIYVHINIKLHILHQVLSQGLRMILQUPJLDLSHNIPSI--- 762

QY 233 SPINKTQGLQGLVLL--IDY--GLRFRVQALSYIACNSIGRFTCKEN----- 276

DB 763 ---NFVSKLHLILEKRPPIIALQEK-VQNILFYFQGLVNEIPAQENSLMTLFIIE 816

QY 277 ---DWCRCCKPFTNCTVS-----MDIQAME--ENLLRITHTKAYNSDPHS-- 320

DB 819 LPFADSKIH-----LHDTGVYSHRIYKKWIMNKSQIQAEDLIRFINIKRQIGKPEHPPG 874

QY 321 -----DPEKIFMKP-----LPMNYELNTSTIMILWTMDSNFPQRYEQENS 361

DB 875 KEIQKLEKILHYKQKVIQYQDVWSEMAKNIILISPSSELSPFLASAPTORILQKHKN 934

QY 362 MKQ--LFAKAKIVHKIPSTSKS---THKAPLISLQKQKINISYVWIKRISLYVNPNDIC 417

DB 935 LTKQLLTHMENAKVAMEDECKKPVAVSK---INLEPHLMNLIILLKLPQFASIGASVLIV 991

QY 418 SFSEEHSHSICENGVAVI---AFLEPTVGLASALITCAPINRTGCTGNTGVMISQGLC 474
 DB 992 C---ENIKLIPDENLITQPSHALI-----SVLLDFPQFQESYSESEITFFHINPPE 1041
 QY 475 KPEVAESTDHYTGFTDQDLEMK-----YLLQ-----KIDPPIEVHA 512
 DB 1042 TSELLVFKIAYGNLAEFLNALKUSLFATADVUNKADFLIQIDFTYKFRUKTNKGKLLHS 1101
 QY 513 IFISNIMRLNSWFDPSWPKPMILIKSNKYKSSI VHMIGLSIGLCTKNTSTELPVLAVY 572
 DB 1102 IV-NEDFKLYEETILTELPK-----KATDAKESLTCLKFASKQKIELSRIT---PLAKET 1151
 QY 573 VN---PFGGSHS---ESWEMPVNENSEPDWE-----RIKLILPLQCYNW 610
 DB 1152 INIEVNFQGVNVVPRPSAFKPAIEVEIKNVPMKINILKPNSHSHIKRID-HIQLIN- 1209
 QY 611 TITIANKKWTKFTFTHIYLPSPKSNPN---GNESIYYEPELEFI---DPSRNLGYM 661
 DB 1210 -LSFDNKKWDFV-----SKSKISFPPELTISSEQLIKQPIKVITETLNKASNDLAVY 1260
 QY 662 -----KINNIQV-----FGYSMHFDPHAIRHILQIDYPTGSSQISALIQ 702
 DB 1261 ISEKILKWLIVFVKELNTEFFVAINMSEFGEVTFDYKHFALEYEINSKYIEIKNIQNE 1420
 QY 703 LLEIRDRVNKLS 714
 DB 1321 ILENTDNIHKLS 1332

RESULT 14

S45734

probable membrane protein YBL004W Yeast (Saccharomyces cerevisiae)

N:A:Protein names: hypothetical protein YBL0101

C:Species: Saccharomyces cerevisiae

C>Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 19-Apr-2002

A:Accession: S45734; S45733; S44556; S37318

R:Rieger, M.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45734

A:Accession: S45734

A:Molecule type: DNA

A:Residues: 1:2066:CRIE

A:Cross-references: EMBL:Z35765; MIPS:YBL004W

A:Experimental source: strain S288C

R:Lohan, A J F; Wolfe, K H.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45730

A:Accession: S45733

A:Molecule type: DNA

A:Residues: 1214-2493:GWSL

A:Cross-references: EMBL:Z35765; MIPS:YBL004W

A:Experimental source: strain S288C

R:Wolfe, K H; Lohan, A J F.

Yeast 10(Suppl.A), S41-S46, 1994

A:Title: Sequence around the open-reading frame of Saccharomyces cerevisiae chromosome II: si

A:Accession number: S44556; M01D:94378721; PMID:8091860

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1214-2493:GWSL

A:Cross-references: EMBL:Z26494; NID:9403311; PID:9403312

A:Experimental source: strain S288C

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 199

C:Genetics:

A:Cross-references: S37318; S44556

A:Map position: 2L

C:Keywords: transmembrane protein

F:165-182/Domain: transmembrane #status predicted <IM1>

F:350-366/Domain: transmembrane #status predicted <IM2>

F:656-672/Domain: transmembrane #status predicted <IM3>

F:1021-1037/Domain: transmembrane #status predicted <IM4>

F:1170-1187/Domain: transmembrane #status predicted <IM5>

Db 464 RLDF 467

Search completed: May 19, 2003, 10:34:40
Job time : 60 secs

Repeat: Transmembrane.
 KW SIGNAL 1 31
 FT PROPEP 32 44
 FT CHAIN 45 587
 FT DOMAIN 56 112
 TSP TYPE-1 1
 FT DOMAIN 117 159
 LDL-RECEPTOR CLASS A.
 FT TRANSMEM 308 324
 POTENTIAL.
 FT TRANSMEM 329 348
 POTENTIAL.
 FT DOMAIN 495 531
 PGP-LIKE
 FT DOMAIN 541 587
 TSP TYPE-1 2.
 FT DISULFID 118 129
 BY SIMILARITY.
 FT DISULFID 123 142
 BY SIMILARITY.
 FT DISULFID 374 399
 BY SIMILARITY.
 SQ SEQUENCE 587 AA: 65540 MW: 6D80417659FC879E CRC64;

Query Match 3.7%; Score 153; DB 1; Length 587;
 Best Local Similarity 20.4%; Pred. No. 0.0011;
 Matches 74; Conservative 57; Mismatches 136; Indels 96; Gaps 16;

QY 35 VSDHATSPD---WLSKRPF---HPSQYTFDFVPSQGFSTRYKI-----YR 79
 DB 201 IQDVRKKPYNQYTLKGTIDFKLOSFSYSSEVHYTMRSSKTTVSIGFALPCVA 260
 QY 80 EFG-----RKVNIAVERNEL-----GSPDLPEPFERNILL 114
 DB 261 EFGFNADSKYSKSEKKIRASRKENSFYQAKAEQLQARYIKKREDLMLHPEFFELRAL 320
 QY 115 GRPPTLQQTENLIKKGTHFLSATLGGESLTIFVDKRLSKRAEGSDSTNSSVTL 174
 DB 321 PQSYNGEVRQ-LYKDYGTHYTEALGGDYEVTVILKEKLEKTYGSLAYKNCQIVL 379
 QY 175 ETUQLAASYFT-----DPSSTLPL-----IHQIAS-----TAIKVTET 210
 DB 380 KVCANIKGVYTVGLGGCGDGLNMGHDIVKGSMDVDYVAVYGGDSFSTWLAANKL 439
 QY 211 PTGFL---GCSNYNLDSSVY-----LVQ-----SPENKLGQLGLVLFYQERFV 256
 DB 440 PTPLRLMRCEAVHYNLDFTRSVTPLYELVTARPESSANSLK-KNCRPALAAYLEE--- 495
 QY 257 QAALSYIACNSGEFTCKENDCWCHGKFPKNCPS-----MDIQAMENILRLTETWKA 312
 DB 496 SSSCRAPCRNNGELAVLKGTPE-----ECVTPSGYSGLGCEITORTDIGDSMSC 546
 QY 313 YNS 315
 DB 547 WGS 549

RESULT 2
 YMW5_YEAST
 ID YMW5_YEAST STANDARD: PRT: 684 AA.
 AC Q04779;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 78.8 kDa protein in ABE2-CHL12 intergenic region.
 OS YMW075W OR YMW916.14.
 ON Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=S288c / A9472;
 RA Pearson D., Bowman S., Harrell H.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 CC 1- SIMILARITY: TO S.POMBE SPAC16C9.05 AND SPAC2F7.07C.
 CC 1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 : entities requires a license agreement (see http://www.ebi.ac.uk/EMBL/seqdb/seqdb.html) or send an email to license@isb-sib.ch).

CC
 CC
 CC
 DR EMBL: Z48952, CAAB8800.1;
 DR SCD: S0004680; YMK075W.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1
 DR SMART: SM00249; PHD; 2.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS0016; ZF_PHD_2; 1.
 KW Hypothetical protein; Zinc-finger.
 FT ZN_FING 260 309 PHD-TYPE.
 FT DOMAIN 15 30 POLY-SER.
 SQ SEQUENCE 684 AA: 78836 MW: CFC282DC9A782F5C CRC64;

Query Match 3.3%; Score 145.5; DB 1; Length 684;
 Best Local Similarity 18.6%; Pred. No. 0.027;
 Matches 128; Conservative 86; Mismatches 221; Indels 256; Gaps 36;

QY 76 KIVRPGRWKV-NNLAVGRNPLGSDPLAP---EPRNRIILGNRPLOOITENLIK 129
 DB 122 K-KRR-SLWNYKKNIGQSNN SEMTIVSKKPTGVRFNFGNENMLKIFLENMIE 177
 QY 130 KYGTHFLSATLGGESLTIFVDKRLSKRAEGSDSTNSSVTLTLHQLAASVETDR 189
 DB 178 ESN-----INSTGWCNDI-----NPTREPESDNDKKLSNIETKILISINAIY---D 225
 QY 190 SYLRRLHHLQIASTAKVTETHTGPGCSNYDNLDSSVIVQSPNPKIQIQLGLVLPD 249
 DB 226 SKSKLFGQNSIKSTS---NASEKIFEDKNS-----TID 256
 QY 250 YLOERFVQAAISYIACNSGPEICKENDCWCHGKFPKNC---PSMDIQAMENILRL 307
 DB 257 FINEHPCS-----ACNSGSPIC-----CDTGKSPHFLCLDPPIDNNLPKQDWHON 364
 QY 308 ETWKAYNSDFEESDEFKLMKRLPMNYETINTSTIMHWTMDSNFQRYEGLNSMKQLFL 367
 DB 305 EC-----KPKIF-----NSMALKKIESNFK-----UNNVKIFA 347
 QY 368 KAKIVHKLFSLSKCHKOPILSLPROKISLYWLIRGSPLYCHNDRGASPSHSIHSTI 427
 DB 338 KL-----LFNDSHNPQ---FOUPNYIKETPAVK-----TGRQYSDK 475
 QY 428 QTNQAVVTPAPITGVQDASACITCAPNRPFGGTFNRYVMSQKFKFVASESHVYD 487
 DB 376 ---NDKIPLT-----DRO---LFTSY---QUSITK-----LDSY-N 492
 QY 488 FETDLODLEMVYLQKTDPTTEVNAIFISNIMPLNSWTFPSWPFMLLTLSNPKYSSLV 547
 DB 403 PDTHDSNSGKFLI-----CYKCNQIRGASWSDHNSRLIM----- 434
 QY 548 HMLGILSIQICLTKKNSLTLEPVAVVNPFGGSHSKSFWP VNSHSPDWERIKLIDLPQ 606
 DB 446 ---GLYCVT-----PHLLAVPRASEKN----- 459
 QY 607 CYNWTLITLGNKKW-----TFFETVH-----YLSRISKNGINGNSIYEP 648
 DB 460 -----LGSKKCKPLHSITKVKYKIHHCORIDNSVYKVKKKKRL-----INKNKLIYEP 508
 QY 649 LEFIDPSRLGYMKINNIVFGYSMHFDPFAIRLIQLDYPYTGQSDSALLQLLEIRD 708
 DB 509 IQ-----KIGYNNCNQIQTPTSHDYD-----FNQDKITQIDENSIKVDFE---D 553
 QY 709 RVNKLSPQQRKLDLFLSHRLKLSISE 738
 DB 554 KIYKSKMKVKKRLFOQESLIDKLVSNGSQ 583

RESULT 3
 CO7_HUMAN
 ID CO7_HUMAN STANDARD: PRT: 844 AA.
 AC P10643; Q92489;

(1) (b) (5) - Exempt (4)
 (2) (b) (5) - Exempt (4)
 (3) (b) (5) - Exempt (4)
 (4) (b) (5) - Exempt (4)
 (5) (b) (5) - Exempt (4)
 (6) (b) (5) - Exempt (4)
 (7) (b) (5) - Exempt (4)
 (8) (b) (5) - Exempt (4)
 (9) (b) (5) - Exempt (4)
 (10) (b) (5) - Exempt (4)
 (11) (b) (5) - Exempt (4)
 (12) (b) (5) - Exempt (4)
 (13) (b) (5) - Exempt (4)
 (14) (b) (5) - Exempt (4)
 (15) (b) (5) - Exempt (4)
 (16) (b) (5) - Exempt (4)
 (17) (b) (5) - Exempt (4)
 (18) (b) (5) - Exempt (4)
 (19) (b) (5) - Exempt (4)
 (20) (b) (5) - Exempt (4)
 (21) (b) (5) - Exempt (4)
 (22) (b) (5) - Exempt (4)
 (23) (b) (5) - Exempt (4)
 (24) (b) (5) - Exempt (4)
 (25) (b) (5) - Exempt (4)
 (26) (b) (5) - Exempt (4)
 (27) (b) (5) - Exempt (4)
 (28) (b) (5) - Exempt (4)
 (29) (b) (5) - Exempt (4)
 (30) (b) (5) - Exempt (4)
 (31) (b) (5) - Exempt (4)
 (32) (b) (5) - Exempt (4)
 (33) (b) (5) - Exempt (4)
 (34) (b) (5) - Exempt (4)
 (35) (b) (5) - Exempt (4)
 (36) (b) (5) - Exempt (4)
 (37) (b) (5) - Exempt (4)
 (38) (b) (5) - Exempt (4)
 (39) (b) (5) - Exempt (4)
 (40) (b) (5) - Exempt (4)
 (41) (b) (5) - Exempt (4)
 (42) (b) (5) - Exempt (4)
 (43) (b) (5) - Exempt (4)
 (44) (b) (5) - Exempt (4)
 (45) (b) (5) - Exempt (4)
 (46) (b) (5) - Exempt (4)
 (47) (b) (5) - Exempt (4)
 (48) (b) (5) - Exempt (4)
 (49) (b) (5) - Exempt (4)
 (50) (b) (5) - Exempt (4)
 (51) (b) (5) - Exempt (4)
 (52) (b) (5) - Exempt (4)
 (53) (b) (5) - Exempt (4)
 (54) (b) (5) - Exempt (4)
 (55) (b) (5) - Exempt (4)
 (56) (b) (5) - Exempt (4)
 (57) (b) (5) - Exempt (4)
 (58) (b) (5) - Exempt (4)
 (59) (b) (5) - Exempt (4)
 (60) (b) (5) - Exempt (4)
 (61) (b) (5) - Exempt (4)
 (62) (b) (5) - Exempt (4)
 (63) (b) (5) - Exempt (4)
 (64) (b) (5) - Exempt (4)
 (65) (b) (5) - Exempt (4)
 (66) (b) (5) - Exempt (4)
 (67) (b) (5) - Exempt (4)
 (68) (b) (5) - Exempt (4)
 (69) (b) (5) - Exempt (4)
 (70) (b) (5) - Exempt (4)
 (71) (b) (5) - Exempt (4)
 (72) (b) (5) - Exempt (4)
 (73) (b) (5) - Exempt (4)
 (74) (b) (5) - Exempt (4)
 (75) (b) (5) - Exempt (4)
 (76) (b) (5) - Exempt (4)
 (77) (b) (5) - Exempt (4)
 (78) (b) (5) - Exempt (4)
 (79) (b) (5) - Exempt (4)
 (80) (b) (5) - Exempt (4)
 (81) (b) (5) - Exempt (4)
 (82) (b) (5) - Exempt (4)
 (83) (b) (5) - Exempt (4)
 (84) (b) (5) - Exempt (4)
 (85) (b) (5) - Exempt (4)
 (86) (b) (5) - Exempt (4)
 (87) (b) (5) - Exempt (4)
 (88) (b) (5) - Exempt (4)
 (89) (b) (5) - Exempt (4)
 (90) (b) (5) - Exempt (4)
 (91) (b) (5) - Exempt (4)
 (92) (b) (5) - Exempt (4)
 (93) (b) (5) - Exempt (4)
 (94) (b) (5) - Exempt (4)
 (95) (b) (5) - Exempt (4)
 (96) (b) (5) - Exempt (4)
 (97) (b) (5) - Exempt (4)
 (98) (b) (5) - Exempt (4)
 (99) (b) (5) - Exempt (4)
 (100) (b) (5) - Exempt (4)

or send an email to license@slb-stb.ch.

CC EMBL: J03507; AAA51861.1;
 CC EMBL: X86328; CAA60121.1;
 DR EMBL: X86328; CAA60121.1; JOINED.
 DR EMBL: X86330; CAA60121.1; JOINED.
 DR EMBL: X86331; CAA60121.1; JOINED.
 DR EMBL: X86332; CAA60121.1; JOINED.
 DR EMBL: X86333; CAA60121.1; JOINED.
 DR EMBL: X86334; CAA60121.1; JOINED.
 DR EMBL: X86335; CAA60121.1; JOINED.
 DR EMBL: X86336; CAA60121.1; JOINED.
 DR EMBL: X86337; CAA60121.1; JOINED.
 DR EMBL: X86338; CAA60121.1; JOINED.
 DR EMBL: X86339; CAA60121.1; JOINED.
 DR EMBL: X86340; CAA60121.1; JOINED.
 DR EMBL: X86341; CAA60121.1; JOINED.
 DR EMBL: X86342; CAA60121.1; JOINED.
 DR EMBL: X86343; CAA60121.1; JOINED.
 DR EMBL: X86344; CAA60121.1; JOINED.
 DR PIR: A27349; A27340.
 DR HSSP: Q07954; ICR8.
 DR Genew: HGNC:1346; C7.
 DR MIM: 217070;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001884; FacI_MAC.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001862; MAC_perforin.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR InterPro: IPR000884; TSPI.
 DR Pfam: PF00057; ldl_recept_a; 2.
 DR Pfam: PF00084; sushi; 4.
 DR Pfam: PF00090; tsp_1; 4.
 DR Pfam: PF01823; MACPF; 2.
 DR PRINTS: PR00764; COMPLEMENTC9.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00057; FIMAC; 2.
 DR SMART: SM00192; ldl_a; 1.
 DR SMART: SM00457; MACPF; 1.
 DR SMART: SM00209; TSPI; 2.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS00068; LDLRA_2; 1.
 DR PROSITE: PS00279; MAC_PERFORIN; 1.
 DR PROSITE: PS00092; TSPI; 2.
 KW Complement pathway; Complement alternate pathway; Glycoprotein;
 KW Plasma; Membrane attack complex; Cytolysis; Signal; EGF-like domain;
 KW Repeat; Sushi; Transmembrane; Disease mutation.
 FT SIGNAL 1 22
 FT CHAIN 23 843
 FT DOMAIN 24 83
 FT DOMAIN 83 121
 FT TRANSMEM 271 287
 FT TRANSMEM 292 311
 FT DOMAIN 451 487
 FT DOMAIN 497 545
 FT DOMAIN 570 627
 FT DOMAIN 630 689
 FT DOMAIN 695 770
 FT DOMAIN 771 843
 FT DISULFID 85 96
 FT DISULFID 91 109
 FT DISULFID 103 119
 FT DISULFID 337 353
 FT DISULFID 571 613
 FT DISULFID 599 626
 FT DISULFID 631 673
 FT DISULFID 659 688
 FT CARBOHYD 46 36
 FT CARBOHYD 292 292
 FT CARBOHYD 503 503
 FT CARBOHYD 506 506
 COMPLEMENT COMPONENT C7
 TSP TYPE 1 1.
 LDL RECEPTOR CLASS A
 POTENTIAL.
 POTENTIAL.
 EGF LIKE.
 TSP TYPE-1 2.
 SUSHI 1.
 SUSHI 2.
 COMPLEMENT CONTROL FACTOR 1 MODULE 1.
 COMPLEMENT CONTROL FACTOR 1 MODULE 2.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 C-LINKED (MAN).
 N-LINKED (GLCNAC...) (POTENTIAL).
 C-LINKED (MAN) (POTENTIAL).
 C-LINKED (MAN) (POTENTIAL).

DB 926 --PQESCLLY-----NR[NH]-----KKRLSGYEC 948

RESULT 6

SR68_YEAST STANDARD; PRT; 599 AA.

AC P36887;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Signal recognition particle 68 kDa protein homolog (SRP68).

GN SRP68 OR YP1243W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID:4932;

KN [1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE

RC STRAIN-S288C;

RC MEDLINE-95009940; PubMed-7425282;

RA Brown J.D., Hann B.G., Medzihradsky K.F., Niwa M., Burlingame A.L.,

RA Walter P.

RT "Subunits of the Saccharomyces cerevisiae signal recognition particle

RT required for its functional expression.";

RL EMBO J. 13:4390-4400(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AR972;

RC MEDLINE-97313271; PubMed-9169875;

RA Russey H., Storms R.K., Ahmed A., Albrecht K., Allen P., Ansoorge W.,

RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Hens V.,

RA Atslein D., Howman S., Bruckner M., Carpenter J., Cherry J.M.,

RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,

RA Dietrich F.S., Delliou H., DiPaolo T., Dubois E., Duesterhoeft A.,

RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goifeau A.,

RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier B.,

RA Hunnicke-Smith S., Hyman R., Johnston M., Kaiman S., Kleine K.,

RA Komp C., Kurdi G., Kashkari D., Lew H., Lin A., Lin D., Louis R.J.,

RA Marathe R., Messenguy F., Mewes H.-W., Miripatis S., Muesel D.,

RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,

RA Pelel F.X., Pohl T.M., Purnelle D., Schaler M., Scharfe M.,

RA Scherrens H., Schramm S., Schroeder M., Sidcu A.M., Tettelin H.,

RA Urrestarazu I.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,

RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,

RA Zhong W.W., Zollner A., Vo D.H., Hani J.

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";

RL Nature 387:103-105(1997).

CC -1- FUNCTION: SIGNAL-RECOGNITION PARTICLE ASSEMBLY HAS A CRUCIAL ROLE

CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC

CC RETICULUM MEMBRANE. SRP68 BINDS THE 7S RNA. SRP72 BINDS TO THIS

CC COMPLEX SUBSEQUENTLY. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT

CC INTERACT DIRECTLY WITH THE DOCKING PROTEIN IN THE PP MEMBRANE AND

CC POSSIBLY PARTICIPATE IN THE ELONGATION ARREST FUNCTION.

CC -1- SUBUNIT: YEAST SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA

CC MOLECULE (SCR1) AND AT LEAST SEVEN PROTEIN SUBUNITS: SRP72, SRP68,

CC SRP54, SRP65, SRP21, SRP14 AND SRP7.

CC -1- SIMILARITY: BELONGS TO THE SRP68 FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: 135177; AAA3401.1;

DR EMBL: 267751; CAA91601.1;

DR EMBL: 273593; CAA97964.1;

DR PIR: S47928; S47928

DR SGD: S0006164; SRP68

KW Signal recognition particle, Ribonucleoprotein, RNA-binding.

SO SEQUENCE 599 AA; 69006 MW; 241749187CE95499 CRC64;

Query Match 2.98; Score 117.5; DB 1; Length 599;

Best Local Similarity 19.98; Pred. No. 0.46;

Matches 72; Conservative 54; Mismatches 120; Indels 115; Gaps 15;

QY 87 NNLAVRRNFFLQSLPLAPDFERNIPRLQRPPLQVLTENLKKYK----- 132

DB 168 NNLAL-----VPAALRHKKNSILIA-----EFVVDNINVKYQYSLKQYAGNLTIP 213

QY 133 --IHFLSATLGG-----ESTLIFDK-----KKLSKRAEGSDSTNS----- 170

DB 214 FNNFIVFVQSPFNKKKPPVWLLDNGENMKFKLIIISLEQKVTININWRSFNKLIIDAF 273

QY 171 -SVILE THGLAASVFIQKSTLRK-----HHIOLASIAIKVITRTQPLGCSNYNL 223

DB 274 VAGFLEQGLSIHTPTQITQYLRSLKLEKALDHIEFFA-----NIDDD 316

QY 224 DSVSSVIVQSPFNKIQGLQVLFHFLQGLAFVVAALSYIAVNSRGEFFICKENICWCHON 283

DB 317 DDIEMVENSNN-----QIIAYIAYNIIISISHERDLFIHW 357

QY 284 PRPEPCNCSMDIQAMENILRITETWKAYNSDFEE-----SDEF--KLEMKRLPMNYF 335

DB 358 NQMLKINTSLPSKILKYKEMPIVKNLIKYSIDIMELGVYSIDELLSQIDUCKLYFOLE 417

QY 336 INTSTIMHIVMDSNPKR-----RYPOLEN-----SRQFLFKACKI-VHKLFSLSK 481

DB 418 ENTGFLSVTYQSKGFYMEALAVYVAYEFLNFISETSELEILLPANLSLSNVSRLQK 477

QY 382 R 382

DB 478 R 478

RESULT 7

IL68_RAT STANDARD; PRT; 918 AA.

AC P40190;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 46, Last annotation update)

DE Interleukin-6 receptor beta chain precursor (IL-6R-beta) (Interleukin

DE 6 signal transducer) (Membrane glycoprotein 130) (gp130).

GN IL6ST.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

EX MEDLINE-93052397; PubMed-1427893;

RA Wang Y., Nesbitt J.E., Fuentes N.L., Fuller G.M.

RT "Molecular cloning and characterization of the rat liver IL-6 signal

KL transducing molecule, gp130.";

KL Genomics 14:666-672(1992).

CC -1- FUNCTION: SIGNAL TRANSDUCING MOLECULE. THE RECEPTOR SYSTEMS FOR

CC IL-6, LIF, OSM, CNTF, AND IL-11 CAN UTILIZE GP130 FOR INITIATING

CC SIGNAL TRANSMISSION. BINDS TO IL-6 IL-6 ALPHA CHAIN) COMPLEX,

CC RESULTING IN THE FORMATION OF HIGH-AFFINITY IL-6 BINDING SITES.

CC AND TRANSDUCES THE SIGNAL. DOES NOT BIND IL-6. MAY HAVE A ROLE IN

CC EMBRYONIC DEVELOPMENT (BY SIMILARITY).

CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- TISSUE SPECIFICITY: FOUND IN HEPATOCYTES, ASTROCYTES, FIBROBLASTS

CC AND ENDOTHELIAL CELLS.

CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN

CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

* Distribution as in 121 as its content is in no way
statement is not removed. Sage by and for commercial
license agreement (See <http://www.isb-sib.ch/annouce/>
to license@isb-sib.ch)

V01 ANNOTATED_CDS

42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

QY 402 TRQSFYCNENCLL-----CSSESETHSC-----PNDVAVVIA 447
Db 382 NGTELIVNLNNRYVASLAARNVVKSPATVLTIPGSHFKASH VYDLKAFPKDLJAVE 441
QY 438 FLDP-----CTVGDASACLTICADNKRIGTINRYM--LSQGLA-- 474
Db 442 WTPPSKPVNKYUUEWCVLSKNSPCT---PWQUCDQVNRHFKGSLKSKYLTIVTPV 498
QY 475 ---KPEVAESTDIYIGFETDLOLEMK-----YLLQKTDRETEVIAFIISNIMRLNS 523
Db 449 FQGGPSPPSMKAYIKCAAPSKGPTVPTKFKVKNIAVIFWGHFVYVNCPTFN 552
QY 524 WDPDSWR---KRMILTAKSNKYKSLVHMILGLSLQICLTKNSTEFV LAVVVNFQ 578
Db 553 -YSISYKTSVGKEMVVRVDSSTHTYTLSSL-----SSCTLYVWUAAAYTEERK 600
QY 579 SHSESFMVYNENSFPDWRKIKIDLPQCYNWII--LGNKKKIFEFVHIVPSIKSN 636
Db 601 DGPEFTFTLL---KFAQGEIAIVPV-GLAFILILILG---VGFENKRFKAKHWP 653
QY 637 GPNESIIYEPLEFIDPSRNLGYMKINNOVEGYSMIIFDEAIFILLIOLYPTQSG 696
Db 654 VP-----DPSKS-----HIAWSPHPP KHNNSKLOUMYSIANF 687
QY 697 DSALLQILFIRKVNKISPPGQRRIOLF-----SOLIRHKLKUSE 748
Db 688 TD--VSVWEIEANKKKPCDGLKSLTLKFKKISTEGHSSII--SSWSSAPPSSISSE 744

RESULT 8

TEGU_HSV7J

ID TEGU_HSV7J STANDARD: PRT: 2059 AA.
AC F52362;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Large tegument protein.
GN 031.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN 11
RP SEQUENCE FROM N.A.
RA Nicholas J
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV 1 UL36,
CC EBV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch)
CC
DR FMRU: 043400; AAC54693.1;
SQ SEQUENCE 2059 AA: 239476 MW: AAE6CA6DEDC5D10 CRC64;

Query Match 2.8%, Score 113, EB 1, Length 2059;
Best Local Similarity 18.9%, Pred No. 5.5;
Matches 160; Conservative 128, Mismatches 298, Indels 216, Gaps 37;

QY 71 FSTRVK YFEFGKWKVNNIAVPRNPFLOSPLAFETRNILGPRDIOGHTNI--- 127
Db 609 FAVLVKLYEDVYSITINGEWEN-----ELIKNI-----EIVDNTQYL 647
QY 128 -IKK*GTHILLSATLGGES IIFVEKPPKSPAPSESTINGCVTETHLQAA 183
Db 648 RQGTENKNTPTTISVFRKFIIEFTSRTIIEPAPFENRFSVIVPGKIAIK 707

FT TRANSMEM 188 204 POTENTIAL.
 FT TRANSMEM 212 231 POTENTIAL.
 FT DOMAIN 373 408 EGF-LIKE.
 FT DOMAIN 416 498 C2 DOMAIN.
 FT DISULFID 257 279 BY SIMILARITY.
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 549 549 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 50 50 V -> M (IN HPLH2).
 FT VARIANT 123 123 /FTID-VAR_010772.
 FT VARIANT 123 123 R -> H.
 FT VARIANT 183 183 /FTID-VAR_010773.
 FT VARIANT 183 183 V -> G (IN HPLH2).
 FT VARIANT 224 224 /FTID-VAR_010744.
 FT VARIANT 224 224 I -> D (IN HPLH2).
 FT VARIANT 225 225 /FTID-VAR_010774.
 FT VARIANT 225 225 R -> W (IN HPLH2).
 FT VARIANT 252 252 /FTID-VAR_010745.
 FT VARIANT 252 252 N -> S (IN HPLH2).
 FT VARIANT 279 279 /FTID-VAR_010746.
 FT VARIANT 279 279 C -> Y (IN HPLH2).
 FT VARIANT 285 285 /FTID-VAR_010747.
 FT VARIANT 285 285 MISSING (IN HPLH2).
 FT VARIANT 345 345 /FTID-VAR_010775.
 FT VARIANT 429 429 I -> L (IN HPLH2).
 FT VARIANT 429 429 G -> E (IN HPLH2).
 FT CONFLICT 332 332 L -> V (IN REF. 2).
 FT CONFLICT 426 426 G -> S (IN REF. 2).
 SQ SEQUENCE 555 AA; 61377 MW; DDFDEODICAR7586R CRC64,
 Query Match 2.7%; Score 112; DB 1; Length 555,
 Best Local Similarity 24.0%; Pred. No. 1.1;
 Matches 73; Conservative 39; Mismatches 114; Indels 78; Gaps 17;
 QY 102 PLAFERNIRIGCRPP-----TLOQITNIIKKYCTHELISATLGG-- 143
 DB 188 PHPPK---KRALGDLPHFNASTOPAYLRLISNYGTHFIRAVELGGKISALTAIRICEL 244
 QY 144 -----EESLTFVDKRLKRAEGSDSTNSSVYLETIHLQAASY---FIDRDS 190
 DB 245 ALRGITNEVEICLVV---KQVNIQIHGCSLSAKACRKKKKHKKMTASHPHQYRKHSS 301
 QY 191 TLRRLHIIQIASTAKVETRTGPGICSNYDINDSVSYVQSPNKTQIQLGLVIL--P 248
 DB 362 EVWGRIHTSINLLRGI---QASPEYSAWNSLPGSLVD-----YTLEFLHVLDSQ 353
 QY 249 DYLGQRFVQAALSYIACNSRGEFICKENKWCRCQP---KPEP--CNC-----PSMDI 296
 DB 354 DPREALRPAISYLTDRA-----RWRCSPRCPGGRQKSPKRPQCCVCHCSAVITQUC 407
 QY 297 QAMEENL LRIT-ETWKAYNSDEEDEF-KLPKRLPMNYFLNTSTIMI---LWTM 347
 DB 408 CPQRCGLAQIEVFTEIGAWGILWETWETATTAIVYKTFPG---QQLPRTSYVWNNNNPLWSV 453
 QY 348 DSNF 351
 DB 464 RLDF 467
 RESULT 11
 ID ITBL_CHICK
 AC P07228;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Integrin beta-1 precursor (CSAT antigen) (JC22 antigen) (RGD-
 DE receptor).
 GN ITGB1.
 OS Gallus gallus (Chicken).
 OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
 OC Archosauria, Aves, Neognathae, Galliformes, Phasianidae; Phasianinae;

OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Embryonic fibroblast;
 RX MEDLINE=86245073; PubMed=3487386;
 RA Tamkun J.W., Desimone D.W., Fonda D., Patel R.S., Buck C.,
 RA Horwitz A.F., Hynes R.O.;
 RT "Structure of integrin, a glycoprotein involved in the transmembrane
 RT linkage between fibronectin and actin";
 RL Cell 46:271-282(1986).
 CC FUNCTION INTEGRINS ALPHA 1/BETA-1, ALPHA 2/BETA-1, ALPHA 3/BETA-1,
 CC 1 AND ALPHA-11/BETA-1 ARE RECEPTORS FOR COLLAGEN. INTEGRINS ALPHA-
 CC 1/BETA-1 AND ALPHA-2/BETA-1 RECOGNIZE THE PROLINE-HYDROXYLATED
 CC SEQUENCE G-P-P-G-E-P IN COLLAGEN. INTEGRINS ALPHA 2/BETA-1, ALPHA-
 CC 3/BETA-1, ALPHA-4/BETA-1, ALPHA-5/BETA-1, ALPHA-6/BETA-1, ALPHA-
 CC 10/BETA-1, ALPHA-11/BETA-1 AND ALPHA-V/BETA-1 ARE RECEPTORS FOR
 CC FIBRONECTIN. ALPHA-4/BETA-1 RECOGNIZES ONE OR MORE DOMAINS WITHIN
 CC THE ALTERNATIVELY SPLICED CS-1 AND CS-5 REGIONS OF FIBRONECTIN.
 CC INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRINOGEN. INTEGRIN
 CC ALPHA-1/BETA-1, ALPHA-2/BETA-1, ALPHA-6/BETA-1 AND ALPHA-7/BETA-1
 CC ARE RECEPTORS FOR LAMININ. INTEGRIN ALPHA-4/BETA-1 IS A RECEPTOR
 CC FOR VCAM1 AND RECOGNIZES THE SEQUENCE Q-1-D-S IN VCAM1. INTEGRIN
 CC ALPHA-9/BETA-1 IS A RECEPTOR FOR VCAM1. CYTOACTIN AND
 CC OSTEOPONTIN. IT RECOGNIZES THE SEQUENCE A-E-I-D-G-I-E-I IN
 CC CYTOACTIN. INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR EPILIGRIN
 CC AND THROMBOSPONDIN. INTEGRIN ALPHA-V/BETA-1 IS A RECEPTOR FOR
 CC VITRONECTIN. BETA-1 INTERFERS WITH THE SEQUENCE R-G-D IN A
 CC WIDE ARRAY OF LIGANDS.
 CC -! SUBUNIT: HEHEKOUIMEK OF AN ALPHA AND A BETA SUBUNIT. BETA-1
 CC ASSOCIATES WITH EITHER ALPHA-1, ALPHA-2, ALPHA-3, ALPHA-4, ALPHA-
 CC 5, ALPHA-6, ALPHA-7, ALPHA-8, ALPHA-9, ALPHA-10, ALPHA-11 OR
 CC ALPHA-V.
 CC -! SUBCELLULAR LOCATION: Type I membrane protein.
 CC -! SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -! SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www1.sib.scrib.ch>)
 CC or send an email to license@sib.scrib.ch.
 CC -----
 CC EMBL, M14049, AAA48926.1, .
 CC PIR: A23947; IJCH3.
 CC HSP: P05106; IJCH3.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR002369; Integrin_H.
 CC InterPro: IPR001169; Integrin_beta_C.
 CC InterPro: IPR003659; Plexin-like
 CC InterPro: IPR002035; VWFA.
 CC Pfam: PF00362; Integrin_B; 1.
 CC PRINTS: PR01186; INTEGRIN.
 CC ProDom: PD001811; Integrin_B; 1.
 CC SMART: SM00187; INB; 1.
 CC SMART: SM00423; PSI; 1.
 CC SMART: SM00327; VWFA; 1.
 CC PROSITE: PS00243; INTEGRIN_BETA; 3.
 CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
 CC Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Phosphorylation; Signal.
 FT SIGNAL 1 24 BY SIMILARITY.
 FT CHAIN 25 803 INTEGRIN BETA-1.
 FT DOMAIN 25 733 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 734 756 POTENTIAL.
 FT DOMAIN 757 803 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 144 382 VWFA-LIKE (POTENTIAL).
 FT DOMAIN 471 640 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 471 520 1.
 FT REPEAT 521 564 11.
 FT REPEAT 565 603 111.


```

QY 202 STAI-----KVETET-----212
Db 1076 QSSWCKKKGQGGFFDFTTIMPKSEFKKQSKKPIVSVEGNNVLAFTNENIHRHPNIP 1135
QY 213 QDQCSNAYND-----SVSSVLQSHENKIQGQIQVQLP-----248
Db 1146 GYLQYITTCNSVDIASLEKLSWSLL-----NORTLEHETDPTIEVNVNCAIVTCGE 1190
QY 249 -DY-----LORRFVQAALSY-----263
Db 1191 GUYCMQDCKFTWFFLPKGLFDGDFKPNPPIPVYVGSKTEERKV ASMAVYKCATSSL 1249
QY 264 --ACNSEGEFTKENIAWHPGKFEFENISMDJAMEENLLAITETWKAYNSDFEESD 321
Db 1250 KAALRAGGVFWAYGJTNANWDALDIAN--TPVQJSEKQIQAIDPIPTSANIT-HRLD 1305
QY 322 EFKLPKKRPLPMNYPLNTSTIMHLWMTDSNFORRYHOLKNSMKQPLKAQKIVHKLPKLSK 381
Db 1306 DGATTIKETPASSAYFSS-----YTHISNDQ-QVLEFVNPPVDSNIIVQQLMITGLGIE 1359
QY 382 RCHKQPLISLPHORTISTYMWLIRQSFVYCNNGILGSHSETHSCICNHOVVCTAFLPC 441
Db 1360 TYNPNPI-----PTSSSEHJ-----LHUTSS-----C 1383
QY 442 TVGDASACTIAPDNPTGNTGNTGMI SGLPYEVAEASHVIRPETIATLPMKYLL 501
Db 1384 CIPPDVCCITICE-----SPSLIPQITVPYINPVYDINPLADYEI-----1423
QY 502 OKTDPPFVHAIPIPSNIMPELNSWFOISWKKRMKILTKSKNYKSSLVHMLIGLSIQICLTK 561
Db 1424 -----AHLDSLVSQARIGSIHYSLQDR--IDLLAHLTAKOMINSIIGLDESVELLN 1473
QY 562 NSTLEHVLAVYNNPQGSSESFWFVNKSNPPOWFKILJLPLQCYNWTITLQ-----615
Db 1474 DAV--VTSIDYN-----NW-----ISGCSY-----TKDILVFKMAWNLJLHLSQMY 1514
QY 616 -----NKWKTFEFTVHLYLSRTKSNPGNESIYEP-----LEFIDFSR-----NL 658
Db 1515 YLRITWTSNIDFYTWTFE-PIPGNALTNIAATLSHPKLI PRANNIDVITVHPAYLASL 1573
QY 659 GYMKINNVQVGYSMHFDPEALPDILGLDYPVYQGSQSAIJLJLHRYVKNKLSPPQ 718
Db 1574 DYKL-SIDAQWGT---KQVLDLUNGITVETILFSETS-----IELSPAMNLI---A 1621
QY 719 RRLDLPSCILRRKL 734
Db 1622 RKLTLALQGNQVL 1637

```

RESULT 13

RESULT: 13	ME25_SCHPO	STANDARD:	PRT: 622 AA.
ID	ME25_SCHPO	STANDARD:	
AC	Q9P6S4; Q96WS0; P78952;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Meiotic expression upregulated protein 25.		
CN	MEU25 OR SHPC27.03.		
OS	Schizosaccharomyces pombe (Fission Ycast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales, Schizosaccharomycetaceae,		
OC	Schizosaccharomycetes.		
OX	NCBI_TaxId=4896;		
RN			
RC	SEQUENCE FROM N. A.		
RC	STRAIN-972;		
EX	MEDLINE-21848401; PubMed-11859360,		
RA	Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,		
RA	Spouras J., Peat N., Hayles J., Baker S., Hasham D., Bowman S.,		
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	Collins M., Connor R., Crawford A., Davis P., Deitweil L., Fraser A.,		
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	Holtroyd S., Hornsby T., Howarth E.J., Hunt S., Jagsels K.,		

[illegible]

DR PROSITE: PS00222: EGF_1: FALSE_NEG.
DR PROSITE: PS01186: EGF_2: FALSE_NEG.
DR PROSITE: PS02079: MAC_PEPFORIN_1:
DR PROSITE: PS04499: C2_DOMAIN_1:
DR PROSITE: PS00004: C2_DOMAIN_2:
KW Glycoprotein, Transmembrane, Cytolysis, Calcium binding, Signal;
KW EGF-like domain; T-cell.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 554 PERFORIN 1.
FT TRANSMEM 187 503 POTENTIAL.
FT TRANSMEM 211 230 POTENTIAL.
FT DOMAIN 372 407 EGF-LIKE.
FT DOMAIN 415 497 C2 DOMAIN.
FT DISULFID 256 278 BY SIMILARITY.
FT CARBOHYD 204 204 N LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 400 400 N LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 548 548 N LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 354 AA: 61513 MW: 784669 EDPAR866F CRK%4;

Query Match: 2.7%; Score 109.5; DH 1; Length 554;

Best Local Similarity 18.8%; Pred. No. 1.7;

Matches 89; Conservative 54; Mismatches 141; Indels 189; Gaps 22;

QY 85 KVNLAVERR NF-----LQSPLEPLAEPRRIEELQK KPILOQL 124
DB 152 KIANFAAEKAGQGYNTDTVFGRMYSPFAKPPHHPFA *NTHNNSSTHAY 211
QY 125 ENLIKKGTHIELLSAIGREEL-----TIFV-----FKLSKEAEISD 164
DB 212 KRLISSGTHPTAVDGGPVSVIALPQLILLDGLIAFVGVGVLSVLAQVSGCAQASV 271
QY 165 STNSSVTLTILHQAASYEIDKDSILKRLHRIQIAIAKVEIRIQLGCSN---VD 221
DB 272 SSEYKAGEKKKQKIATSF-----HOTYRERH-----VEVLGGELSSNLLRG 316
QY 222 NIDS-----VSSVLYQSPENKIQGLQVILPDY--LQFREVQAALSYIA-----C 265
DB 317 NQATPEHSTWIASLPTKPVVYVSLEPHILLESPPKPKREALKJALSHVYSKRWKRC 376
QY 266 N SENEFTKENLWHSQPKPEPNPSMDIQAMEENLLKLETKWYKNSDFEESD 322
DB 377 NRPCRAQHKSSRSDSCVC-----QDSNVINQCCPRQGLAKI-----C 416
QY 323 FKLPKRLPMWNYPLNTSTIMHLWTMDSNQRR YQLLNKMKQLFKKQKIVHKLP 377
DB 417 -----MYRNFQAKGLWDYITSTIAYLKVFEGQEI -----447
QY 378 SLKRCFKQPLISLPKQRISTYWLTRIQSF-----LYCNGNG---414
DB 448 -----RTGVVNNNHPSSDKMDKFNVLISGGEPKRVGVVWIAWGWD 490
QY 415 -LLGGSFEET---HSCICPD-----QVVCYAFPCYVHSAALTCAP 454
DB 491 DLIGTCRKSPKSGFHEVNCPLNHGSKFIYQANC---LPIELGEP LCLEYAP 548

RESULT 15

ID DYHC SCUPO STANDARD; PRT: 4196 AA.
AC G13250; Q9P3L0; Q90T18;
BI 15-JUL-1998 (Ref. 36, Created)
DI 15-JUN-1998 (Ref. 36, Last sequence update)
DI 15-JUN-2002 (Ref. 41, Last annotation update)
DE Dysein heavy chain, cytosolic (DYHC).
GN D1C1 OR SPN130C2.31C OR SPAC1093.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota,ungi, Ascomycota, Schizosaccharomycetes.
OC Schizosaccharomycetales, Schizosaccharomycetaceae.
OC Schizosaccharomyces.
OX NCBI_TaxID=1896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CML132;

DR PROSITE: PS00222: EGF_1: FALSE_NEG.
DR PROSITE: PS01186: EGF_2: FALSE_NEG.
DR PROSITE: PS02079: MAC_PEPFORIN_1:
DR PROSITE: PS04499: C2_DOMAIN_1:
DR PROSITE: PS00004: C2_DOMAIN_2:
KW Glycoprotein, Transmembrane, Cytolysis, Calcium binding, Signal;
KW EGF-like domain; T-cell.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 554 PERFORIN 1.
FT TRANSMEM 187 503 POTENTIAL.
FT TRANSMEM 211 230 POTENTIAL.
FT DOMAIN 372 407 EGF-LIKE.
FT DOMAIN 415 497 C2 DOMAIN.
FT DISULFID 256 278 BY SIMILARITY.
FT CARBOHYD 204 204 N LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 400 400 N LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 548 548 N LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 354 AA: 61513 MW: 784669 EDPAR866F CRK%4;

Query Match: 2.7%; Score 109.5; DH 1; Length 554;

Best Local Similarity 18.8%; Pred. No. 1.7;

Matches 89; Conservative 54; Mismatches 141; Indels 189; Gaps 22;

QY 85 KVNLAVERR NF-----LQSPLEPLAEPRRIEELQK KPILOQL 124
DB 152 KIANFAAEKAGQGYNTDTVFGRMYSPFAKPPHHPFA *NTHNNSSTHAY 211
QY 125 ENLIKKGTHIELLSAIGREEL-----TIFV-----FKLSKEAEISD 164
DB 212 KRLISSGTHPTAVDGGPVSVIALPQLILLDGLIAFVGVGVLSVLAQVSGCAQASV 271
QY 165 STNSSVTLTILHQAASYEIDKDSILKRLHRIQIAIAKVEIRIQLGCSN---VD 221
DB 272 SSEYKAGEKKKQKIATSF-----HOTYRERH-----VEVLGGELSSNLLRG 316
QY 222 NIDS-----VSSVLYQSPENKIQGLQVILPDY--LQFREVQAALSYIA-----C 265
DB 317 NQATPEHSTWIASLPTKPVVYVSLEPHILLESPPKPKREALKJALSHVYSKRWKRC 376
QY 266 N SENEFTKENLWHSQPKPEPNPSMDIQAMEENLLKLETKWYKNSDFEESD 322
DB 377 NRPCRAQHKSSRSDSCVC-----QDSNVINQCCPRQGLAKI-----C 416
QY 323 FKLPKRLPMWNYPLNTSTIMHLWTMDSNQRR YQLLNKMKQLFKKQKIVHKLP 377
DB 417 -----MYRNFQAKGLWDYITSTIAYLKVFEGQEI -----447
QY 378 SLKRCFKQPLISLPKQRISTYWLTRIQSF-----LYCNGNG---414
DB 448 -----RTGVVNNNHPSSDKMDKFNVLISGGEPKRVGVVWIAWGWD 490
QY 415 -LLGGSFEET---HSCICPD-----QVVCYAFPCYVHSAALTCAP 454
DB 491 DLIGTCRKSPKSGFHEVNCPLNHGSKFIYQANC---LPIELGEP LCLEYAP 548

RESULT 15

ID DYHC SCUPO STANDARD; PRT: 4196 AA.
AC G13250; Q9P3L0; Q90T18;
BI 15-JUL-1998 (Ref. 36, Created)
DI 15-JUN-1998 (Ref. 36, Last sequence update)
DI 15-JUN-2002 (Ref. 41, Last annotation update)
DE Dysein heavy chain, cytosolic (DYHC).
GN D1C1 OR SPN130C2.31C OR SPAC1093.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota,ungi, Ascomycota, Schizosaccharomycetes.
OC Schizosaccharomycetales, Schizosaccharomycetaceae.
OC Schizosaccharomyces.
OX NCBI_TaxID=1896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CML132;

Yamamoto A., West P.R., McIntosh J.P., Hiraoka Y.,
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

[2]
SEQUENCE FROM N.A.

STRAIN-972;

MEDLINE-21848401; PubMed-11859460;

Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
Mooney K., Moule S., Mundell K., Murphy I., Nibbelin D., Odell C.,
Olivey K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
Rutherford K., Puttler S., Saunders D., Seeger K., Sharp S.,
Skellion J., Simmonds M., Squares R., Squares K., Stevens K.,
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
Woodward J., Volkov G., Ait P., Robben J., Grymonprez B.,
Wolffens J., Vansteels E., Kieger M., Schaefer M., MacIver Auer S.,
Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
Eger P., Zimmermann W., Wedler H., Wambuit R., Purnelle B.,
Goffeau A., Cadieu E., Drenth S., Gloux S., Lelaure V., Mottier S.,
Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
Lucas M., Kohert M., Galland J., Tallada V.A., Garza A., Thode C.,
Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
Domínguez A., Revuelta I., Moreno S., Armstrong J., Forsburg S.L.,
Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Polashkin J.,
Shpakovski G.V., Ussery D., Harrell B.G., Nurse P.,
"The genome sequence of *Schizosaccharomyces pombe*,"
Nature 415:871-880(2002).

-1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
ORGANELLES ALONG MICROTUBULES. REQUIRED FOR NUCLEAR MOVEMENT
DURING MEIOTIC PROPHASE.

-1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
INTERMEDIATE AND LIGHT CHAINS.

-1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROBABLY BINDS INDIRECTLY TO
THE INNER PLASMA MEMBRANE.

-1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL: AB006784; BAA22056.1;

EMBL: AL355652; CAB90788.1;

EMBL: AL132839; CAB60251.1;

KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.

FT DOMAIN 1217 1252 COILED COIL (POTENTIAL).

FT DOMAIN 1984 2012 MICROTUBULE-BINDING (POTENTIAL).

FT DOMAIN 3315 3403 COILED COIL (POTENTIAL).

FT DOMAIN 3643 3666 COILED COIL (POTENTIAL).

FT NP-BIND 1890 1897 ATP (POTENTIAL).

FT NP-BIND 2169 2176 ATP (POTENTIAL).

FT NP-BIND 2174 2181 ATP (POTENTIAL).

FT NP-BIND 2520 2527 ATP (POTENTIAL).

SQ SEQUENCE 4196 AA: 484-408 MW: 861046370184600 CAC664;

Query Match 2.7%; Score 109; DB 1: Length 4196,

Best local Similarity 19.6%; Pred. No. 29;

Matches 165; Conservative 123; Mismatches 286; Indels 268; Gaps 46.

QY 85 KVNIAVERNRFLGSLPLAEPFRNI-PL---LAPRPTLQOITENLIKKGTHILLSAT 140

Db 274 QISKTALEENFW------QFYRSLRLNQLHSRPVFLVD---ILAFGRFHTAS 323

QY 141 LGGHESITLFDVK-RKLSKRAEGSDSTNNSSSVTLTQHIAASYFLDRDS----- 190

DB 324 ENSEINIOCFVDKVKCKDALEKEISLIDIFSSSTLESLSLSAALLSYSTSKKWNTOYF 382

QY 191 TLPLRLHHIQ-TASTAIKVTEPTPTGPGSNYENLD-----SVSSVL----- 240

DB 383 ETRVLDIFNF-TREDILKILSRLLPALGALSNDVDSHRTAVSSDLSLOYRIKIDFLRI 442

QY 231 ---VQSPENKIQOGIQVLLDPTVQFRFVQAALSYLACNSE--G 269

DB 443 SGLSKEEGSYGLKNSIKVIRAFENKLLK-----YIQS-FHEKHOQLIGALSEVYG 491

QY 270 PFICKENDCWHGCGPKPFNCFCGM-DIGAMENMLRIET---ETWKA-----YNSDFE 318

DB 492 LTHITELFELHUNKKEHVFNLTVFKDIQSL--NVLDSIKGVNANSLSTSYNCMTV 549

QY 319 ESEDFELFKMLPMKYFLNTSTIMHWTMSNCFPPY--EGLNSMKO-LFLKAKIVHK 375

DB 550 LDRVFAIQLSL-LQY-----SKTSSOMFTIMKQFPTFRVTRTISDCLHIANRTRK 605

QY 376 LFSLSKRCHK-----QPLSIPKORISIVMLIKUSLFY---CNENCLIGS----- 418

DB 606 LDKLTRFEDVSDELIAAMELRLNPASSAIIWATOLKRLHEYTKNINIEGEDWNN 665

QY 419 -----FSHETHSCDPCPNQV 433

DB 666 FIDGFEKVECIHQKRTDNIIFTNWVSSPNINFEFESKTEFYLTQSSAESPLRS 725

QY 434 VCTAFPTVGTASACLTCA-----PLNTRDQICNIGYMLSGGLCKPEVAESTHYIG 487

DB 726 VSDIDHPSF--CKRIPTLAHGYNIPQSLMEHASCQRIQLI-AMCLIDEVQSPND VS 781

QY 488 FETDQQLKEMKYLQNTDRIEVH---AIFIS-NDMRINSWFDP----- 527

DB 782 FEISKTE-EERFELQEYELAVRQHIVTGLTISWDFVGNLSIPEK-ALGKRNFKNHP 840

QY 528 -SWRKRMLITLKSNNKYSSIVHMLIGI-----SLQIC-----LTKNSTLEPVL- 569

DB 841 NVENYAYQFSSTLSLIMNK RNATSHYMQIQEGLPQADICEYSGDLELIDRKLQDLID 899

QY 570 AVYNPFGGSHSESWFMPVNSFDFWERTKLDLPLOCYNNWTLILGNKWKTEFTFVHLYL 629

DB 900 LIYVNGY-----SNLPPFVRA---LNLNPDILLISKCKPLSFPKLT----- 948

QY 630 ASRIKSNCFNGNE-----SIYYEPLFIDFSRNIQYMAINNIOVGYSMHFDPAIR-- 681

DB 949 ---ILTSGNENNDLKSKFSSDMYEKLK-----GFLKPTNLTIOPTNIEFDPVVEKK 987

QY 682 ---DLILQ IDYIV TQGSQCSALQDILHFKRWVKNLSFGGGRRLHFSCLL 728

DB 988 EDSIVLLDMCIQSVDNIPDLISIKITTAQRNCTLGHFPVY---INKLE---SLHIGTFPSLI 1041

QY 729 RH 730

DB 1042 FH 1043

Search completed: May 19, 2003, 10:32:02
Job time : 40 secs

before you hit the 4578
step (v) 1993 2003 1993 1993

our the using sw m del

20 3, 10, 27, 18 : Search time 97 seconds
(without alignments)
(with alignments)
(with alignments)
(with alignments)

332 129-35

SRGAHPSLMAFW

ANXANTMTMDYITLCS 766

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

332 129-35

17 122 3.0 1723 5 Q8W8D0
18 121 3.0 967 5 Q8W8D0
19 119.5 2.9 785 5 Q8W8D0
20 118 2.9 1165 2 Q45844
21 118 2.9 1165 2 Q45887
22 117.5 2.9 990 2 Q8F004
23 117 2.9 617 2 Q8F004
24 116.5 2.9 945 10 Q8F004
25 116.5 2.9 1299 17 Q8F004
26 115.5 2.8 940 2 Q8F004
27 115 2.8 617 2 Q8F004
28 115 2.8 967 5 Q8F004
29 114.5 2.8 504 5 Q8F004
30 114 2.8 504 5 Q8F004
31 114 2.8 6885 4 Q8W8D0
32 113.5 2.8 454 10 Q8W8D0
33 113.5 2.8 458 10 Q8W8D0
34 113 2.8 2559 12 Q8W8D0
35 112 2.7 2346 5 Q8W8D0
36 112 2.7 2346 5 Q8W8D0
37 111.5 2.7 658 12 Q8W8D0
38 111 2.7 514 5 Q8W8D0
39 111 2.7 1489 3 Q8W8D0
40 110.5 2.7 770 10 Q8W8D0
41 110.5 2.7 1830 16 Q8W8D0
42 110 2.7 464 5 Q8W8D0
43 110 2.7 622 3 Q8W8D0
44 110 2.7 1072 6 Q8W8D0
45 110 2.7 1141 17 Q8W8D0

ALIGNMENTS

RESULT 1

Q95560 ID Q95560 PRELIMINARY: PRI: 781 AA.
AC Q95560
DT 01-MAY-1999 (FEMBLrel. 10, Created)
DT 01-MAY-1999 (FEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (FEMBLrel. 17, Last annotation update)
DE Hypothetical 88.7 kDa protein (Unknown) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:7606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (JAN-1979) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 576-781 FROM N.A.
RX MEDLINE:9627227; PubMed:8519474;
RA Andersson B., Wentland M.A., Ricciarelli J.Y., Liu W., Gibbs R.A.;
RT "A 'double adapter' method for improved shotgun library
construction";
RL Anal. Biochem. 236:107-113(1996).
RN [3]
RP SEQUENCE OF 576-781 FROM N.A.
RX MEDLINE:97264341; PubMed:9110174;
RA Yu W., Andersson B., Wentland M.A., Ricciarelli J.Y., Liu W.,
PA Ricciarelli J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT "Large scale concatenation cDNA sequence library";
RL Genome Res. 7:353-358(1997).
RN [4]
RP SEQUENCE OF 576-781 FROM N.A.
RA Mei G., Yu W., Gibbs R.A.;
RT Submitted (FEB 1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL: A1035789; CAA22893.1;
DR EMBL: A131933; AAD20055.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001179; FKBP_PPase.
DR InterPro: IPR002384; GLA_bone.

DR InterPro: IPR001862; MAC_perforin.
 DR PRINTS: PR00002; GLABONE.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00457; MACPF; 1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 781 AA; DEE3EB83A08B599 CRC64;

Query Match 72.6%, Score 2962, DB 4, Length 791,
 Best Local Similarity 70.6%, Pred No Gaps 242;
 Matches 543; Conservative 99; Mismatches 109; Indels 18; Gaps 4;

QY 16 ALWEH---TALSILHCWLVAAV-----VSDQAT-----SPFDWLLSKGPPHRSQETDF 63
 DB 13 AVAPMTALLALGPGWLVAVSATAAAYVPPROHASVAGOHPLDMLITUNGPPHRAKAYADF 72
 QY 64 VLSKSGSTSTPKYIPETIPKVNNAIAVERPNFIASPLAPAEFPRNIPILGPRPTLQGI 124
 DB 73 MEYRQGFITRYPIYREFAFAPKVNNAIAEPKPFESILPLAPAEFPRNIPILGPRPNLQV 142
 QY 124 TENLIKKYGTHTPLSATIGGRESLTIIPVDKKLSKRAKGSSTT-----NSSSVILETL 177
 DB 133 TENLIKKYGTHTPLSATIGGRESLTIIPVDKQKLGKRTTTCGASLIGSGNSTAVSLETL 192
 QY 178 HQLAASYFIDPESTLPPHHTGATGATKVTETPTPLDPSNINLDSVSVLVQSPENK 247
 DB 193 HQLAASYFIDPESTLPPHHTGATGATKVTETPTPLDPSNINLDSVSVLVQSPENK 252
 QY 238 IQGLQGLVLLPVLGPRVQALSYIACNSGPPHCKPKNCWCHGCKPKPPPCNSMDIQ 297
 DB 253 VQLGLQVLLPVLGPRVQALSYIACNSGPPHCKPKNCWCHGCKPKPPPCNSMDIQ 312
 QY 298 AMEENLRITETWKAAYSDFEEDFKLEMKRKLPMNYFLNTSTIMHLWTMDSNFORRYEQ 357
 DB 313 AMEDSLIQDSWATHNQFESEEFQALIKRIPDRPLNSTAISQFWAMTSLQHRVQ 372
 QY 358 LENSMLQLFLKAKIVHKLFSLSKRCHKOPILSIPLROTSYTWLTIQSLFYCNENGLIG 417
 DB 373 LGAGLKVLFKKTHTPLPFLNFNCKRCHPQPPFLKPEPSLYWNNAIQLSLYNSSTEPG 432
 QY 418 SFSEETHSCOTPNQOVCTATPLCTVIGDASACTTCADPNTRCTGCTGYMISGLCKPE 477
 DB 433 TFEUSHSCTCPYDSSCCGPIPCALGEGACACATNSTRCSNCPGVYLAAGLCRPE 492
 QY 478 VAESTDHYIGFTDIDQEMKYLLQKTDPTFVHAFTISNDMLNSWPDPSWRKRMILLT 537
 DB 493 VAESLENFLGLETUQLEIKYLLQKQSPTEVHSIFISNDMLPLGSMWPDPSWRKRMILLT 552
 QY 538 KSNKYKSSLVHMTLGLSLQICLTKNSTLPIVAVVNPFGGSHSESWFMPVNSFPDWE 597
 DB 553 KSNKYKPIGVHVMIALSLQICLTKNSTLPIVMAIVVNPFGGSHSESWFMPVNSFPDWE 612
 QY 598 RTKLDPQLQKWNITLGNKWKIFETVHLYLRSIKSNIPGNESITYEPELEFIDPSRN 657
 DB 613 RTWVDAQAQCNWTITLGNRKWTFETVHLYLRSIKSLDSSNETIYIEPELEMTDPSKN 672
 QY 658 LGYMKINNTQVGVSMHFLPEALPOLLQLLYPYLQSSQALLQLELPRVKNLSPPG 717
 DB 673 LGYMKINNTQVGVSMHFLPEALPOLLQLLYPYLQSSQALLQLELPRVKNLSPPG 732
 QY 718 QRRLDFSLRLHRLKLTSEVVRISQALQAFNAKIPMTIMVITKIFS 766
 DB 733 KVRLLDLPSCILRHRIKIANNEVGRIOSSIRAFNSKLPNPVHYETCKICS 781

RESULT 2
 Q9C0B6

ID Q9C0B6 PROTEINAFY, ppt, 791 AA

AC Q9C0B6

DT 01-JUN-2001 (Trembl et al., 17, Created)

DT 01-JUN-2001 (Trembl et al., 17, Last sequence update)

DT 01-JUN-2001 (Trembl et al., 18, Last annotation update)

DE KIAA1747 protein (Fragment).
 GN KIAA1747.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN 1
 RP SEQUENCE FROM N.A.
 PX MEDLINE=21082922; PubMed=11214970;
 RA Nagase T, Kikuno F, Hattori A, Kondo Y, Okumura K, Ohara O.
 RT Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.
 PL JNA Res. 7:347-356(2000).
 DR FBL: AB051534; BAB21838.1;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001179; FKBP_PPIase.
 DR InverPro: IPR001862; MAC_perforin.
 DP SMART: SM00181; EGF; 1.
 DP SMART: SM00457; MACPF; 1.
 DP PROSITE: PS00453; FKBP_PPIASE_1; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 791 AA; 89887 MW; 21C5CD35AAFC6A89 CRC64;

Query Match 72.6%, Score 2962, DB 4, Length 791,
 Best Local Similarity 70.6%, Pred No Gaps 242;
 Matches 543; Conservative 99; Mismatches 109; Indels 18; Gaps 4;

QY 16 ALWEH---TALSILHCWLVAAV-----VSDQAT-----SPFDWLLSKGPPHRSQETDF 63
 DB 23 AVAPMTALLALGPGWLVAVSATAAAYVPPROHASVAGOHPLDMLITUNGPPHRAKAYADF 82
 QY 64 VLSKSGSTSTPKYIPETIPKVNNAIAVERPNFIASPLAPAEFPRNIPILGPRPTLQGI 123
 DB 73 MEYRQGFITRYPIYREFAFAPKVNNAIAEPKPFESILPLAPAEFPRNIPILGPRPNLQV 142
 QY 124 TENLIKKYGTHTPLSATIGGRESLTIIPVDKKLSKRAKGSSTT-----NSSSVILETL 177
 DB 133 TENLIKKYGTHTPLSATIGGRESLTIIPVDKQKLGKRTTTCGASLIGSGNSTAVSLETL 202
 QY 178 HQLAASYFIDPESTLPPHHTGATGATKVTETPTPLDPSNINLDSVSVLVQSPENK 247
 DB 203 HQLAASYFIDPESTLPPHHTGATGATKVTETPTPLDPSNINLDSVSVLVQSPENK 262
 QY 238 IQGLQGLVLLPVLGPRVQALSYIACNSGPPHCKPKNCWCHGCKPKPPPCNSMDIQ 297
 DB 253 VQLGLQVLLPVLGPRVQALSYIACNSGPPHCKPKNCWCHGCKPKPPPCNSMDIQ 322
 QY 298 AMEENLRITETWKAAYSDFEEDFKLEMKRKLPMNYFLNTSTIMHLWTMDSNFORRYEQ 357
 DB 323 AMEDSLIQDSWATHNQFESEEFQALIKRIPDRPLNSTAISQFWAMTSLQHRVQ 382
 QY 358 LENSMLQLFLKAKIVHKLFSLSKRCHKOPILSIPLROTSYTWLTIQSLFYCNENGLIG 417
 DB 383 LGAGLKVLFKKTHTPLPFLNFNCKRCHPQPPFLKPEPSLYWNNAIQLSLYNSSTEPG 442
 QY 418 SFSEETHSCOTPNQOVCTATPLCTVIGDASACTTCADPNTRCTGCTGYMISGLCKPE 477
 DB 443 TFEUSHSCTCPYDSSCCGPIPCALGEGACACATNSTRCSNCPGVYLAAGLCRPE 502
 QY 478 VAESTDHYIGFTDIDQEMKYLLQKTDPTFVHAFTISNDMLNSWPDPSWRKRMILLT 537
 DB 503 VAESLENFLGLETUQLEIKYLLQKQSPTEVHSIFISNDMLPLGSMWPDPSWRKRMILLT 562
 QY 538 KSNKYKSSLVHMTLGLSLQICLTKNSTLPIVAVVNPFGGSHSESWFMPVNSFPDWE 597
 DB 563 KSNKYKPIGVHVMIALSLQICLTKNSTLPIVMAIVVNPFGGSHSESWFMPVNSFPDWE 622
 QY 598 RTKLDPQLQKWNITLGNKWKIFETVHLYLRSIKSNIPGNESITYEPELEFIDPSRN 657
 DB 623 RTWVDAQAQCNWTITLGNRKWTFETVHLYLRSIKSLDSSNETIYIEPELEMTDPSKN 682
 QY 658 LGYMKINNTQVGVSMHFLPEALPOLLQLLYPYLQSSQALLQLELPRVKNLSPPG 717


```

Db 348 ATTAQKQKIQRTARKLFGISVRCRHNPHQIPRRTIGQWLAHVQSLYCNENGFHWGIFL 407
QY 421 EETHSTCPNDQVQVCTAFPLCTVGDASACLTGAPDNRTROGTGTCNTGYMLSGCLCKPEVAE 480
Db 408 ESQSCVWVHGSLTQKPLPCTVGDASACLTGAPDNRTROGTGTCNTGYMLSGCLCKPEVAE 467
QY 481 S--TQRYIGFETDL--QULEMKYLLQKTDKRVHVAIFISNIMPLNSWFDPSWKKPMILIT 536
Db 468 SERSEVFSEFETDLQDLELYLLQKMDSPLYVITTFISNIPPLITFTFPPWPKPMSLT 527
QY 517 LKSNKYKSLVIMLIGLSQIGLTKNSTLEPVLAVYVNPFGSGHSSESWMFVNPNSEPDW 596
Db 528 LKSNKRNMDP IHMVIGMSKRIQUMRNSSDPMPFVVVNPFGSGHSSESWMFVNPNSEPDW 587
QY 597 ERTKLDLPQCYNMWLTGKWKITFEFTVHLYLSPKISNIPNINFEISYIYPLELELPSR 656
Db 588 EKIRLQ--NSQCYNMWLTGKWKITFEFTVHLYLSPKISNIPNINFEISYIYPLELELPSR 644
QY 657 NIQYMKINNIQVFGYSMHFDPKALRDLIGLQVYPTQGSQ----DSALLQLLELRKVNK 712
Db 645 RQVYKISDVQVFGYSLRFNADLLPSAVQVQVNSYTGSGQFYSSSSVLLMLDLIRDRNR 704
QY 713 LSP---PGQPRIDLSLPHLPKLTSEVWPLQALQAFNAKLNTMDYDTTKLC 765
Db 705 LAPPVAPCKPQDLPSCKMLKHKRLKINSEIRVNHALDLYNTEITLKQSDQMTAKIC 760

RESULT 5
Q925T8
ID Q925T8 PRELIMINARY: PRT: 760 AA.
AC Q925T8
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE BMP/retinoic acid-inducible neural-specific protein.
GN BKNP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Patton;
OX NCBI_TaxID=10116;
RN [1]
RA SOURCE FROM N.A.
RA Matsuda i., Nakatani T., Kawano H., Kobayashi M., Abe A., Araki N.,
RA Asari M.;
RT "Identification and characterization of RPNP, a BMP/retinoic acid
RT inducible neural-specific protein."
RL Submitted (Nov 2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051356; BAB55642.1;
DR SOURCE 760 AA; 88654 MW; 2243964946C26503 CRC64;

Query Match 52.4%; Score 2136.5; DB 11; Length 760;
Best Local Similarity 51.7%; Pred. No. 6.8e-172;
Matches 402; Conservative 133; Mismatches 213; Indels 29; Gaps 10;

QY 1 MWKSRAGAFLSMALMEWIALSLHCWVLAVAAVSDHATSPDMLLSUKGPHRSQF 59
Db 1 MNWRF---VELLYFLVWGRISVQPSH----QEPAGTQGHVSKEPDLWLSDRGPHHSRS 53
QY 60 YTFDEVSFGSGFTPKYKTYPEFSPKVNNAVERPNFLSGPLPLAPEFPPNIRILLGRPT 119
Db 54 YLSFVERHHQGFTRKYKTYREAPKVRNTATERRDLVRHVPVLMPPQRSIRILGRPT 113
QY 120 LQGITENLKKYGTHTPLLSATIGRESLITFVKKKLSKKAEGSDSTINSSVILETLHQ 179
Db 114 TQGFDTITKKYGTHTLLSATIGGEALTYMDKSR-----DRKSGNATQVEALLQ 166
QY 180 LAASYFTDRNSTPLRHHQIQTASTAKVTETPTGPGCSNYDNLSVSSVLQSPENKIO 239
Db 167 LASSYFVDRNGTMRPLHETQISTGALKVTFTPTGLGCSNDNLDSVSSVLQSTESKLI 226
QY 240 LQGITVLLPDYQERTVQAAVLSIACNSEGEFLKCNQWVHGGKPEFECNCPMSDIQAM 299
Db 227 LQGITVLLPDYQERTVQAAVLSIACNSEGEFLKCNQWVHGGKPEFECNCPMSDIQAM 286

```

```

QY 360 PLNLRLELPLWAAVNSDRESDPKLPKRLMAYELNLSLTHMLWMLSGNEQRYTQDE 359
Db 287 EFTLMAKAVAWTEATYKQDLENSEDEFSKPKLPSPNIFTTPTSTHSHWNTQWPLQVPELQ 446
QY 360 NSMKOLPLKAAKILVHKLPLSLKSRCHKQPLISLPHORTSYWLTROSPLHYCNENGLGSP 419
Db 347 SATEAKQKQKIQRTARKLFGISVRCRHNPHQIPRRTIGQWLAHVQSLYCNENGFHWGIFL 406
QY 420 SEETHSTCPNDQVQVCTAFPLCTVGDASACLTGAPDNRTROGTGTCNTGYMLSGCLCKPEVA 479
Db 467 ESQSCVWVHGSLTQKPLPCTVGDASACLTGAPDNRTROGTGTCNTGYMLSGCLCKPEVA 466
QY 480 S--TQRYIGFETDL--QULEMKYLLQKTDKRVHVAIFISNIMPLNSWFDPSWKKPMILIT 535
Db 467 SERSEVFSEFETDLQDLELYLLQKMDSPLYVITTFISNIPPLITFTFPPWPKPMSLT 526
QY 517 LKSNKYKSLVIMLIGLSQIGLTKNSTLEPVLAVYVNPFGSGHSSESWMFVNPNSEPDW 595
Db 528 LKSNKRNMDP IHMVIGMSKRIQUMRNSSDPMPFVVVNPFGSGHSSESWMFVNPNSEPDW 586
QY 596 ERTKLDLPQCYNMWLTGKWKITFEFTVHLYLSPKISNIPNINFEISYIYPLELELPSR 655
Db 587 EKIRLQ--NSQCYNMWLTGKWKITFEFTVHLYLSPKISNIPNINFEISYIYPLELELPSR 643
QY 656 NIQYMKINNIQVFGYSMHFDPKALRDLIGLQVYPTQGSQ----DSALLQLLELRKVN 711
Db 644 RQVYKISDVQVFGYSLRFNADLLPSAVQVQVNSYTGSGQFYSSSSVLLMLDLIRDRNR 704
QY 712 LSP---PGQPRIDLSLPHLPKLTSEVWPLQALQAFNAKLNTMDYDTTKLC 765
Db 704 LAPPVAPCKPQDLPSCKMLKHKRLKINSEIRVNHALDLYNTEITLKQSDQMTAKIC 760

RESULT 6
Q90X10
ID Q90X10 PRELIMINARY: PRT: 761 AA.
AC Q90X10;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE DHCCR1.
GN DHCCR1 OR DHCCR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA SOURCE FROM N.A.
RA STRAIN=C57; TISSUE=BRAIN;
RA Ochal L.K., Sowden M.P., Messing E.M., Wheelless L.L., Reeder J.E.;
RA "Mouse Dhccr1 cDNA."
RL Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF202896; AAF17579.1;
DR MGI; MGI:1928478; Dhccr1.
DR InterPro; IPR001862; MAC_perforin.
DR SMART; SM00457; MACPF; 1.
DR SEQUNCE 761 AA; 88750 MW; F2792FA84FB9879B CRC64;

Query Match 52.1%; Score 2128; DB 11; Length 761;
Best Local Similarity 51.7%; Pred. No. 3.6e-171;
Matches 402; Conservative 131; Mismatches 216; Indels 28; Gaps 10;

QY 1 MTWSPAFVAFLESLMALMEWIALSLHCWVLAVAAVSDHATSPDMLLSUKGPHRSQF 60
Db 1 MNWRF---VELLYFLVWGRISVQPSH----QEPAGTQGHVSKEPDLWLSDRGPHHSRS 54
QY 61 YTFDEVSFGSGFTPKYKTYPEFSPKVNNAVERPNFLSGPLPLAPEFPPNIRILLGRPT 120
Db 55 YLSFVERHHQGFTRKYKTYREAPKVRNTATERRDLVRHVPVLMPPQRSIRILGRPT 114
QY 121 LQGITENLKKYGTHTPLLSATIGRESLITFVKKKLSKKAEGSDSTINSSVILETLHQ 180

```



```

QY 504 TDRREIVHAIFISNOMPLNSWFDPSWPKRMMLTLTKSNK 541
DB 181 TDRREIVHAIFISNOMPLNSWFDPSWPKRMMLTLTKSNK 218

RESULT 9
Q8WU22 PRELIMINARY: PRT: 320 AA.
AC Q8WU22:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to deleted in bladder cancer chromosome 7 region candidate 1.
DE (130898A) (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Straussberg R.;
RL Submitted (JAN-2002) to +b: EMBL/GenBank/DBJ databases.
DR EMBL: BC021560; AAI21560.1;
DR InterPro: IPR001862; MAC_perforin.
DR SMART: SM00457; MACPF: 1.
SQ SEQUENCE 320 AA: 37168 MW: 216A4CH33764P5G1 CRG64.

Query Match 23.8%; Score 970.5; DB 4; Length 326.
Best Local Similarity 56.4%; Pred. No. 9.6e-74;
Matches 181; Conservative 53; Mismatches 72; Indels 15; Gaps 4;

QY 1 MWRKAGAEFLPSMALWEWIALS-LHCWVLAAVAASDGHATSPFDWLLSDKGFPHRSQK 59
DB 1 MNRRF---VELLYFLFWGRISVQPSH----ORPACTDQHSKKEFDWLLSDKGFPHRSK 53
QY 60 YTFDEVSROGCTPYKYIYPERFQWKNVLAVERPNFISPIAPERPNRLIGRPT 119
DB 54 YLSEVHRHOGCTPYKYIYPERFQWKNVLAVERPNFISPIAPERPNRLIGRPT 113
QY 120 LQQTENILKKYGTHTLLSATIGGHSFSLIFVKKRKLKPAQSDSTNSSSVTLTTHQ 179
DB 114 TQGFIDTILKKYGTHTLLSATIGGHSFSLIFVKKRKLKPAQSDSTNSSSVTLTTHQ 166
QY 180 LAASYFDINRSTPLRIHHQIATAIKVTFTFTPTDPIGGSNYINLDSVSVLVQSPENKIQ 239
DB 167 LASSYFDVDRKLGIMKKLRLHQLISGAIKVTFTFTPTDPIGGSNYINLDSVSVLVQSPENKIQ 226
QY 240 LQGLVLLPDYLGPRFVQAAISYIAGNSEJELDKENLQWILGPKFPEINPSMDIQA 299
DB 227 LQGLVLLPDYLGPRFVQAAISYIAGNSEJELDKENLQWILGPKFPEINPSMDIQA 286
QY 300 PENILRTITETWKAYNDSDFERS 320
DB 287 EYTLANNAKSWAEAYKDLNLS 307

RESULT 10
Q8WX56 PRELIMINARY: PRT: 228 AA.
AC Q8WX56:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to deleted in bladder cancer chromosome 7 region candidate 1.
DE (130898A) (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Babbade A.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL138894; CAD13302.1;
DR InterPro: IPR001862; MAC_perforin.
DR SMART: SM00457; MACPF: 1.
FT NON_TER 228
SQ SEQUENCE 228 AA: 25585 MW: 187CB869E49177414 CRG64.

Query Match 17.7%; Score 720.5; DB 4; Length 228.
Best Local Similarity 57.9%; Pred. No. 8.5e-53;
Matches 140; Conservative 40; Mismatches 47; Indels 15; Gaps 4;

QY 1 MWRKAGAEFLPSMALWEWIALS-LHCWVLAAVAASDGHATSPFDWLLSDKGFPHRSQK 59
DB 1 MNRRF---VELLYFLFWGRISVQPSH----ORPACTDQHSKKEFDWLLSDKGFPHRSK 53
QY 60 YTFDEVSROGCTPYKYIYPERFQWKNVLAVERPNFISPIAPERPNRLIGRPT 119
DB 54 YLSEVHRHOGCTPYKYIYPERFQWKNVLAVERPNFISPIAPERPNRLIGRPT 113
QY 120 LQQTENILKKYGTHTLLSATIGGHSFSLIFVKKRKLKPAQSDSTNSSSVTLTTHQ 179
DB 114 TQGFIDTILKKYGTHTLLSATIGGHSFSLIFVKKRKLKPAQSDSTNSSSVTLTTHQ 166
QY 180 LAASYFDINRSTPLRIHHQIATAIKVTFTFTPTDPIGGSNYINLDSVSVLVQSPENKIQ 239
DB 167 LASSYFDVDRKLGIMKKLRLHQLISGAIKVTFTFTPTDPIGGSNYINLDSVSVLVQSPENKIQ 226
QY 240 LQ 241
DB 227 LQ 228

RESULT 11
Q920P4 PRELIMINARY: PRT: 72 AA.
AC Q920P4:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE BMP/retinoic acid inducible neural specific protein (BRINP)
DE (fragment).
GN BRINP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EMBRYONIC STEM CELLS;
RA Matsumoto K., Ariga H.;
RA Matsumoto K., Nakatani T., Kawano H., Arai N., Abe A., Kobayashi M.;
RT "Identification of BMP/RA-inducible neural specific protein.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB060587; BAB70600.1;
FT NON_TER 72
SQ SEQUENCE 72 AA: 8838 MW: 62CDF1B03256A82D CRG64.

Query Match 4.3%; Score 175; DB 11; Length 72.
Best Local Similarity 46.2%; Pred. No. 2.4e-07;
Matches 36; Conservative 11; Mismatches 25; Indels 4; Gaps 2;

QY 1 MWRKAGAEFLPSMALWEWIALS-LHCWVLAAVAASDGHATSPFDWLLSDKGFPHRSQK 60
DB 1 MNRRF---VELLYFLFWGRISVQPSH----ORPACTDQHSKKEFDWLLSDKGFPHRSK 54
QY 61 TDFVDRSGSGSTRYKIY 78
DB 55 LSPVHRHOGCTPYKYIY 72

RESULT 12
Q9QJ37

```



```

QY 461 GTCNTGCM:SGGICNCFEVALSDHYDGFEDLQLEMKYLLQKIDRRKLEVHALFISNMVK 520
DB 296 STC...GLMK...-----NKLEKFF-----QSSGHVSVRTDGG 324
QY 521 LNSWFDSPSKRMILTLKSNKYKSSVIMILGLSLQJCLTKNSTLEPVLAVVYVNPFGGSH 580
DB 325 AVNNFMP---KDIATHPKLPCKETNIVSVLUGHGCHIF-----K 359
QY 581 SESKMPVNVENSEPDEPTKLDLPQYNNWTLTJLNKKWKTFE---TVHLYLR-----SR 632
DB 360 SKEKCMIGTSTI-NWERN-----TMSNHI-----GNKAVTIGEENDVHLHPYTKENKLST 409
QY 633 IKSNGPNCNSESIVYGFLEPFDPSRNIGYMKINNIQVFG 670
DB 410 IYVKMPSGNEQ-----BELNKATREGPAEENSWKQEG 441

RESULT 14
QY 09TUQ3 PRELIMINARY: PRT: 843 AA.
AC 09TUQ3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Complement component C7 precursor.
GN C7.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.; SEQUENCE OF 640-651, AND TISSUE SPECIFICITY.
RP TISSUE=LUNG;
RX MEDLINE=20341712; PubMed=10878384;
RA Agha A., Montalto M.C., Kiesecker C.L., Morrissey M., Grover M.,
RA Whoolery K.L., Rother R.P., Stahl G.L.;
RT Isolation, characterization, and cloning of porcine complement
RT component C7.
RL J. Immunol. 165:1059-1065(2000).
CC 1- FUNCTION: C7 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX. C7
CC BINDS TO C5B FORMIN; THE C5B-7 COMPLEX, WHERE IT SERVES AS A
CC MEMBRANE ANCHOR.
CC 1- SUBUNIT: MONOMER OR DIMER; AS A C5B-7 COMPLEX IT CAN ALSO FORM
CC MULTIMERIC ROSETTES (BY SIMILARITY).
CC 1- TISSUE SPECIFICITY: BONE MARROW, HEART, INTESTINE, LUNG, SPLEEN,
CC KIDNEY, LIVER AND THYMUS.
CC 1- PTM: C7 HAS 28 DISULFIDE BRIDGES (BY SIMILARITY).
CC 1- SIMILARITY: TO COMPLEMENT FACTORS C6, C8, C9, AND TO PERFORIN.
CC 1- SIMILARITY: CONTAINS 2 SUSHI (SCR) REPEATS.
CC 1- SIMILARITY: CONTAINS 2 TYPE-1 TSP REPEATS.
CC 1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
EMRL: AF162274, AAD47918 1.
DR HSP: Q07954; ICR8.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003884; FACI-MAC.
DR InterPro: IPR002172; LDL_recept.A.
DR InterPro: IPR001862; MAC_perforin.
DR InterPro: IPR000436; Sush1_SCP_CCP.
DR InterPro: IPR000884; TSP1.
DR pfam: PF00057; ldl_recept_a; 1.
DR pfam: PF01823; MACPF; 1.
DR Pfam: PF00084; sush1; 2.
DR Pfam: PF00090; tsp_1; 2.
DR Pfam: PF00764; COMPLEMENTC9.
DR PRINTS: SM00032; CCP; 2.
DR SMART: SM00057; FIMAC; 2.
DR SMART: SM00192; LDLa; 1.
DR SMART: SM00457; MACPF; 1.
DR SMART: SM00209; TSP1; 2.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01209; LDLA_1; 1.

```

```

DR PROSITE: PS00279; MAC_PERFORIN; 1.
DE PROSITE: PS00692; TSP1; 2.
KW Complement pathway; Complement alternate pathway; Glycoprotein;
KW Plasma; Membrane attack complex; Cytolysis; Signal; EGF like domain;
KW Repeat; Sushi; Transmembrane.
PT SIGNAL 1 22 HY SIMILARITY.
FT CHAIN 23 843 COMPLEMENT COMPONENT C7 (BY SIMILARITY).
FT REPEAT 24 83 TYPE-1 TSP 1.
FT DOMAIN 83 121 LDL-RECEPTOR CLASS A.
FT TRANSMEM 271 287 POTENTIAL.
FT DOMAIN 292 311 POTENTIAL.
FT DOMAIN 451 487 EGF-LIKE.
FT REPEAT 497 545 TYPE-1 TSP 2.
FT DOMAIN 570 682 2 X SUSHI (SCR) REPEATS.
FT REPEAT 570 627 SUSHI 1.
FT REPEAT 630 689 SUSHI 2.
FT DOMAIN 771 843 COMPLEMENT CONTROL FACTOR 1 MODULE.
FT DISULFID 85 96 HY SIMILARITY.
FT DISULFID 91 109 BY SIMILARITY.
FT DISULFID 103 119 BY SIMILARITY.
FT DISULFID 337 353 BY SIMILARITY.
FT DISULFID 571 613 HY SIMILARITY.
FT DISULFID 599 626 HY SIMILARITY.
FT DISULFID 631 673 BY SIMILARITY.
FT DISULFID 659 688 BY SIMILARITY.
FT CARBOHYD 702 754 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 843 AA; 93061 MW; 06703094C4F50C4D CPO64;

Query Match 3.18; Score 127; DB 6; Length 843;
Best Local Similarity 20.5%; Pred No. 0.1;
Matches 100; Conservative 56; Mismatches 153; Indels 176; Gaps 23;

QY 55 HRSQYTD-----FVDRSGCFSTRYKLYREFGKWKVNI-----AVRRNTFGSP--- 100
DB 210 HFTSTHTSSSGRVFIFSSSSSSSYAKTYFIIKKSYQILVYONTVVRVAFINNPEP 269
QY 101 LPLAFETFRNPIJGFPPTLQGIT--ENLKKGTHLLSATLGEESLTIFYDKRLKLSK 158
DB 270 LQLAISFWKLSYL---PPIYDYSAYRRLIDYGYTHYLOSGLGGYKVIIFYVDSKVKAR 326
QY 159 RRGSDNTTNSSTVTFETIHLAASFYFDIDSTLPHLHHIQIASTAIK-----VIEITCP 214
DB 327 SDLGSEDKKKGASSHISPLFKSKHKCKAMEALKSASGTQ--SNVLRGVPEVRGGRPGF 384
QY 215 LQCSNYDNLQSVSSVIVGSPENKIQIAGQVIVHVDYTLQERFVCAALSYIACNSKGRPFCK 274
DB 385 VSGI.SYLELDN-----PCKRQRY----- 403
QY 275 ENDCWCHGCPKTECNCFSDIOAMEENLRITETWKAYNSDFEESDEPKLFMKRLPMNY 434
DB 404 --SSWAGSVTDLP-----QVTKOKITPLYE-----LVKEVPC--- 433
QY 335 FLNTSTIMHLWTMDSNFORKYEULENSMKQLPILKAKILVHKLFSLSKRCHQKPLISLPRO 394
DB 434 -----ASVKRLYLK--RALEEYLDLDFDSCHCQP----- 459
QY 395 RIETIYMLTRIQSFLYCNHNGLLGSFSETHSCPC---PNDQVVCVTAFLPC---LVCDAS 447
DB 460 -----C-ONG--GNASVEGTCQCHCKPN-----TFCVACRGGVIVGDHA 496
QY 448 -----ACLT-----CAPDNTKTCGTNTGYMLSQG-LCKPEVAESTDHYIGFETDLQDL 495
DB 437 GGIAGGWSWSSWGPGCAUGKKTRSKCNPNPPSGGKSCIGETSESK-----QCDEFDL 550
; QY 496 EMKYLILQ 502
DB 551 EHLRLLE 557
RESULT 15

```


MINARY: PRI: 2: AA
Score: 12, Q (created)
Score: 12, Last sequence update)
Score: 20, Last update (date)
Date:
6.
virus, no RNA stage: herpesviridae;
Roseolovirus.
9) The EMBL/GenBank/ Databases.
GenBank: 10482554
10) Nakano K., Katagiri M., Mori Y.,
Saito K., Sashihara T., Hara A., Zou P., Kosuge H.,
Complete DNA sequence of human herpesvirus 6
[1999].
AA78252.1:
5' MYB DNA binding
MYB 1: UNKNOWN
A: 249211 MW: 71 kDa 476,444 bp CR664;
4.08; Score 124; OR 12; Length 2077;
20.88; Prod. No. 67;
110; Mismatches 248; Indels 292; Gaps 44;
JGSTRKYIVREFGRKVNLA VERNELGSLPLAPEFERN 111
SKF --KIVD---NNNLAATMIVLINCIPAYL---VNIYNI 633
LQUTENI-----IKYIIIF --SALIGGHSIIIVDKRI 156
ELAEVKIYINTUFTIFTSNINNTSPTLSIFMEVDFYNSQFP 693
SLSINSSSVIIIEGAAAEVETKSLIKRHIIQIASTAKV 207
SIAAPSKNIIVNEM KSLPELCKMEISTENKQ 742
---LGSNYNLSVSS VLVSPF-----NKIQL 240
LITTYDIFRHSAYNINIVKNHRELEKRPDIAAIIHDKIQ- 801
RFV-DAALSVIATNSDEPTVENWVWVGHKPPPCNCPSMDIOAM 299
LITVPEJTESTVL-----FLV -DLFFN---STKTAI 841
AYNSDFESDEKIFMKKIFMNYFINIIMHLWIMDSNFORRYEO 357
KANIKITFEED---LAPILNS-----EQ 877
KIVKESLSKPPHKLFLSLIPPTSYWLTQISLYCENGLLG 417
KKKICIMDRHKEIYSCKITIAEKN WIOR----- 916
NOVVIAPIPTVIGASATATANEHIGICNTGYMSQGLCKPE 477
EE-ETAEF---ATAPKHALQ---TKPE 947
DGLTIFMKVILQIDREKVGAEEENDMKINSWFDPSWKRMEL 535
MEFQMKUJANKKKKNPNLEKRLND-----ILL 987
SLVHMLGLSLQITTKNSTLEPV 569

Search completed: May 19, 2003, 10:33:41
Job time: 105 secs

